



Methods for producing diterpenes

Hamberger, Björn; Møller, Birger Lindberg; Ranberg, Johan Andersen; Bohlmann, Carl Jorg; Zerbe, Philip; Nielsen, Morten Thrane

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Hamberger, B., Møller, B. L., Ranberg, J. A., Bohlmann, C. J., Zerbe, P., & Nielsen, M. T. (2015). Methods for producing diterpenes. (Patent No. WO2015113570).

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



(51) International Patent Classification:

C12P 5/00 (2006.01) C07K 14/415 (2006.01)
C12P 7/00 (2006.01) C12N 9/00 (2006.01)

(21) International Application Number:

PCT/DK2015/050021

(22) International Filing Date:

30 January 2015 (30.01.2015)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

PA 2014 00056 31 January 2014 (31.01.2014) DK

(71) Applicants: UNIVERSITY OF COPENHAGEN

[DK/DK]; Nørregade 10, DK-1165 Copenhagen K (DK).
DANMARKS TEKNISKE UNIVERSITET [—/DK];
Anker Engelundsvej 101 A, 2800 Kgs. Lyngby (DK).

(72) Inventors: HAMBERGER, Björn; Borgbygård Allee 11,

DK-2770 Kastrup (DK). MØLLER, Birger Lindberg;
Kongstedvej 5, DK-2700 Brønshøj (DK). ANDER-
SEN-RANBERG, Johan; Hothers Plads 13, 3.th., DK-
2200 Copenhagen N (DK). BOHLMANN, Carl Jörg;
3706 West 29th Stret, Vancouver, BC, V6S 1T4 (CA).

ZERBE, Philipp; Suite 205, 3732 Mount Seymour Park-
way, North Vancouver, BC, V7G 1C3 (CA). NIELSEN,
Morten Thrane; Heinesgade 10, 1. sal, DK-2200 Copen-
hagen N (DK).

(74) Agent: HØIBERG A/S; St. Kongensgade 59 A, 1264
Copenhagen K (DK).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG,
MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM,
PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC,
SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,
TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,
TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,
DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,

[Continued on next page]

(54) Title: METHODS FOR PRODUCING DITERPENES

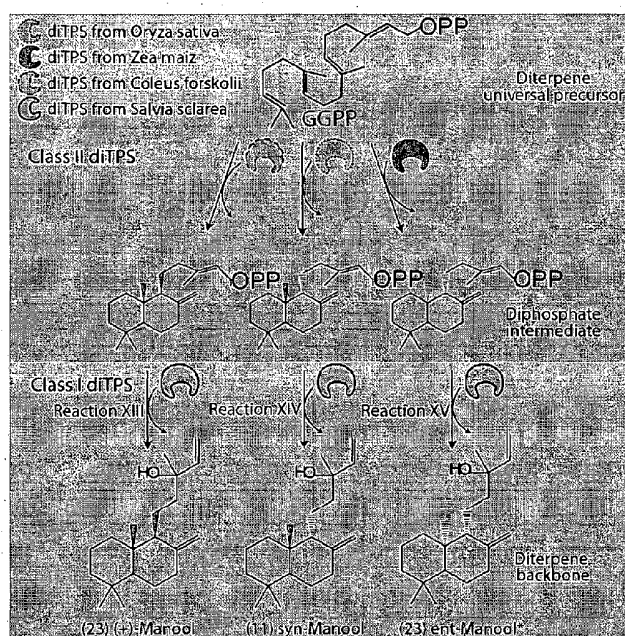


Fig. 1

(57) Abstract: The present invention discloses that by combining different di TPS enzymes of class I and class II different diterpenes may be produced including diterpenes not identified in nature. Surprisingly it is revealed that a di TPS enzyme of class I of one species may be combined with a di TPS enzyme of class II from a different species, resulting in a high diversity of diterpenes, which can be produced.



LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, KM, ML, MR, NE, SN, TD, TG).

Published:

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

— with international search report (Art. 21(3))

— before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments (Rule 48.2(h))

Methods for producing diterpenes

Field of invention

- 5 The present invention relates to the field of biosynthetic methods for producing diterpenes.

Background of invention

- 10 Terpenes constitute a large and diverse class of organic compounds produced by a variety of plants as well as other species. Terpenes modified by oxidation or rearrangements are generally referred to as terpenoids.

Terpenes and terpenoids find multiple uses, for example as flavor compounds, additives for food, as fragrances and in medical treatment

- 15 Terpenes are derived biosynthetically from units of isoprene, which has the molecular formula C_5H_8 . Diterpenes are composed of *four isoprene* units and in nature they are produced from geranylgeranyl pyrophosphate.

Summary of invention

- 20 In nature diterpenes are produced with the aid of specific pairs of diterpene synthases (diTPS) derived from two classes, class I and class II.

- 25 The present invention discloses that by combining different diTPS enzymes of class I and class II different diterpenes may be produced including diterpenes not identified in nature. Surprisingly it is revealed that a diTPS enzyme of class I of one species may be combined with a diTPS enzyme of class II from a different species, resulting in a high diversity of diterpenes, which can be produced.

- 30 Thus, the invention features an inventory of functional class II and class I diTPS from a range of plants, which are useful for accumulating high-value and bioactive diterpenes. When these diTPS are paired into specific modules consisting of new-to-nature combinations, such as using enzymes from different plant species, both the structure and the stereochemistry of the formed diterpenes can be controlled. This strategy gives

access to a novel structural diversity of highly complex diterpenes, representing potentially bioactive molecules, starting materials for chemical synthesis, and intermediates for further functionalization to flavours, fragrances, pharmaceuticals and fine chemicals.

5

The invention thus in one aspect provides methods of producing a terpene, said methods comprising the steps of:

- a) providing a host organism comprising
- 10 I. A heterologous nucleic acid encoding a diTPS of class II,
 II. A heterologous nucleic acid encoding a diTPS of class I,
- with the proviso that said diTPS of class II and said diTPS of class I is not from the same species;
- b) Incubating said host organism in the presence of geranylgeranyl pyrophosphate
- 15 (GGPP) under conditions allowing growth of said host organism;
- c) Optionally isolating diterpene from the host organism.

The invention further provides host organisms, comprising

- 20 I. A heterologous nucleic acid encoding a diTPS of class II;
 II. A heterologous nucleic acid encoding a diTPS of class I,
- with the proviso that said diTPS of class II and said diTPS of class I is not from the same species.

Said host organism may for example be any of the host organisms described herein below in the section "Host organism".

25

It is preferred that the combination of diTPS of class II and diTPS of class I is not found in nature. Thus, it is preferred that the diTPS of class II and the diTPS of class I is not from the same species. Accordingly, if the diTPS of class I is from species X or highly similar to a diTPS of class I of species X, then it is preferred that the diTPS of class II

30 does not have a sequence identity of more than 95%, such as of more than 90%, for example of more than 80%, such as of more than 70% to any diTPS of class II of species X. Similarly, if the diTPS of class II is from species X or highly similar to a diTPS of class II of species X, then it is preferred that the diTPS of class I does not

have a sequence identity of more than 95%, such as of more than 90%, for example of more than 80%, such as of more than 70% to any diTPS of class I of species X. In this connection the term "highly similar" means sharing more than 95%, such as of more than 90%, for example of more than 80%, such as of more than 70% sequence
5 identity.

The invention also provides several enzymes useful with the methods of the invention. Thus, the invention provides EpTPS7 like diTPS enzymes, such as EpTPS7 of SEQ ID NO:2 or a functional homologue thereof sharing at least 70%, such as at least 80%,
10 such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.

The invention also provides TwTPS7 like diTPS enzymes, such as TwTPS7 of SEQ ID NO:4 or a functional homologue thereof sharing at least 70%, such as at least 80%,
15 such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.

The invention also provides CfTPS1 like diTPS enzymes, such as CfTPS1 of SEQ ID NO:5 or a functional homologue thereof sharing at least 70%, such as at least 80%,
20 such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.

The invention also provides TwTPS21 like diTPS enzymes, such as TwTPS21 of SEQ ID NO:7 or a functional homologue thereof sharing at least 70%, such as at least 80%,
25 such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.

The invention also provides TwTPS14/28 like diTPS enzymes, such as TwTPS14/28 of SEQ ID NO:8 or a functional homologue thereof sharing at least 70%, such as at least 80%,
30 such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.

The invention also provides EpTPS8 like diTPS enzymes, such as EpTPS8 of SEQ ID NO:9 or a functional homologue thereof sharing at least 70%, such as at least 80%,

such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.

5 The invention also provides EpTPS23 like diTPS enzymes, such as EpTPS23 of SEQ ID NO:10 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.

10 The invention also provides TwTPS2 like enzymes, such as TwTPS2 of SEQ ID NO:14 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.

15 The invention also provides EpTPS1 like enzymes, such as EpTPS1 of SEQ ID NO:15 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.

The invention also provides CfTPS14, such as CfTPS14 of SEQ ID NO:16 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.

20

Description of Drawings

Figure 1 provides an example of biosynthesis pathways to diterpenes of different stereochemistry. The figure shows biosynthesis of three different isomers of manool by using diTPS enzymes from four different species: *Oryza Sativa* (rice), *Zea maiz* (maize), *Coleus forskolii* (medicinal plant) and *Salvia sclarea* (medicinal plant). The diTPS from *Oryza sativa* may for example be the enzyme of SEQ ID NO:1. The diTPS from *Zea maiz* may for example be the enzyme of SEQ ID NO:3. The diTPS from *Coleus forskolii* may for example be the enzyme of SEQ ID NO:5. The diTPS from *Salvia sclarea* may for example be the enzyme of SEQ ID NO:11.

35 Figures 2A and 2B shows "Combinatorial wheels" showing examples of compounds, which can be made by combining different diTPS enzymes. The universal precursor, GGPP is shown in the middle. The next ring shows various examples of diTPS class II enzymes. The next ring shows various examples of diTPS class I enzymes. The outer ring shows the diterpenes produced by the indicated combinations of diTPS class II

and diTPS class I enzymes. Each diterpene has been assigned a compound number used to identify said diterpene herein. The sequences of all of diTPS class II and diTPS class I enzymes are provided herein in the sequence listing and MS spectras of all the diterpene compounds are given in figure 6. Table 1 also provides a list of the
5 diterpenes.

Figures 3A and 3B show the reactions catalysed by various class II diTPS enzymes as well as the diterpene pyrophosphate intermediates generated by the reactions.

10 Figure 4 shows an alignment of the amino acid sequences of selected diTPS enzymes of class I.

Figure 5 shows an alignment of the amino acid sequences of selected diTPS enzymes of class II.
15

Figure 6 shows MS spectras of hexane extracts from *N. benthamiana* expressing the different diTPS genes. MS spectras of all 47 diterpenes produced as described in Example 1 are shown, with the compound number indicated in the upper left corner of each spectrum. For some compounds also reference spectra are shown.
20

Detailed description of the invention

Method for producing diterpenes

25 The present invention relates to a biosynthetic method for producing diterpenes. The methods typically involves the steps of

- a) Contacting GGPP with a diTPS of class II, which may be any of diTPS of class II described herein in any of the sections "diTPS of class II", "syn-CPP type diTPS", "ent-CPP type diTPS", "(+)-CPP type diTPS", "LPP type diTPS", and "LPP like type diTPS", thereby producing a diterpene pyrophosphate intermediate;
30
- b) Contacting said diterpene pyrophosphate intermediate with a diTPS of class I, which may be any of diTPS of class I described herein in any of the sections
35 "diTPS of class I", "EpTPS8", "EpTPS23", "SsSCS", "CfTPS3", "CfTPS4",

“MvTPS5”, “TwTPS2”, “EpTPS1” , and “CfTPS14” thereby producing a diterpene.

5 It is generally preferred that the diTPS of class I and the diTPS of class II are not from the same species. Furthermore, it is preferred that when said diTPS of class II is SsLPPS then said diTPS of class I is preferably not CfTPS3, CfTPS4 or EpTPS8 and when said diTPS of class I is EpTPS8, then the diTPS of class II is preferably not CfTPS2 or SsLPPS. In particular, when said diTPS of class II is SsLPPS or any of the functional homologues of SsLPPS described in the section “LPP type diTPS”, then said
10 diTPS of class I is preferably not CfTPS3 or any of the functional homologues thereof described in the section “CfTPS3”, is also preferably not CfTPS4 or any of the functional homologues thereof described in the section “CfTPS4”, and is also preferably not EpTPS8 or any of the functional homologues thereof described in the section EpTPS8. It is also preferred that when said diTPS of class I is EpTPS8 or any
15 of the functional homologues thereof described in the section “EpTPS8”, then the diTPS of class II is preferably not CfTPS2 or any of the functional homologues thereof described in the section “LPP type diTPS” or SsLPPS or any of the functional homologues thereof described in the section “LPP type diTPS”.

20 The method may be performed in vitro or in vivo.

The diterpene pyrophosphate intermediate and the diterpene may for example be any of the compounds described herein below in the sections “Diterpene pyrophosphate intermediates” and “Diterpenes”.

25 When the methods are performed in vitro, the above-mentioned steps a) and b) may be performed individually in the indicated sequence, or they may be performed simultaneously. When both steps are performed simultaneously GGPP and the diTPS of class II and the diTPS of class I may all be incubated in the same container under
30 conditions allowing activity of both the diTPS of class II and the diTPS of class I. When the steps are performed sequentially, the step a) may be performed first in one container, whereafter the diTPS of class I may be added to the container. It is also possible that the diterpene pyrophosphate intermediate may be purified or partly purified after step a) and then it may be contacted with the diTPS of class I e.g. in
35 another container.

When the methods are performed in vitro they may contain the steps of providing a host organism comprising

- 5 a. A heterologous nucleic acid encoding a diTPS of class II, which may be any of diTPS of class II described herein in any of the sections "diTPS of class II", "syn-CPP type diTPS", "ent-CPP type diTPS", "(+)-CPP type diTPS", "LPP type diTPS", and "LPP like type diTPS"; and/or
- 10 b. A heterologous nucleic acid encoding a diTPS of class I, which may be any of diTPS of class I described herein in any of the sections "diTPS of class I", "EpTPS8", "EpTPS23", "SsSCS", "CfTPS3", "CfTPS4", "MvTPS5", "TwTPS2", "EpTPS1", and "CfTPS14";
- b) preparing an extract of said host organism;
- c) providing GGPP
- d) incubating said extract with GGPP
- 15 thereby producing a diterpene.

When the methods are performed in vitro they may also contain the steps of

- 20 a) providing a host organism comprising a heterologous nucleic acid encoding a diTPS of class II, which may be any of diTPS of class II described herein in any of the sections "diTPS of class II", "syn-CPP type diTPS", "ent-CPP type diTPS", "(+)-CPP type diTPS", "LPP type diTPS", and "LPP like type diTPS"; and
- b) Preparing an extract of said host organism
- c) Providing another host organism comprising a heterologous nucleic acid
- 25 encoding a diTPS of class I, which may be any of diTPS of class I described herein in any of the sections "diTPS of class I", "EpTPS8", "EpTPS23", "SsSCS", "CfTPS3", "CfTPS4", "MvTPS5", "TwTPS2", "EpTPS1", and "CfTPS14";
- d) preparing an extract of the host organism of c); and
- 30 e) providing GGPP
- f) incubating the extract of step b) and the extract of d) with GGPP OR incubating the extract of b) with GGPP followed by incubating the product with the extract of d)
- thereby producing a diterpene.
- 35

In a preferred embodiment of the invention the methods are performed in vivo. The term "in vivo" as used herein refers that the method is performed within a host organism, which for example may be any of the host organisms described herein below in the section "Host organism". In embodiments of the invention wherein the methods are performed in vivo, it is preferred that steps a) and b) are performed simultaneously. Thus, the methods may comprise the steps of

I. Providing a host organism comprising

- a. A heterologous nucleic acid encoding a diTPS of class II, which may be any of diTPS of class II described herein in any of the sections "diTPS of class II", "syn-CPP type diTPS", "ent-CPP type diTPS", "(+)-CPP type diTPS", "LPP type diTPS", and "LPP like type diTPS",
- b. A heterologous nucleic acid encoding a diTPS of class I, which may be any of diTPS of class I described herein in any of the sections "diTPS of class I", "EpTPS8", "EpTPS23", "SsSCS", "CfTPS3", "CfTPS4", "MvTPS5", "TwTPS2", "EpTPS1", and "CfTPS14"

II. Incubating said host organism in the presence of GGPP under conditions allowing growth of said host organism

III. Optionally isolating the diterpene from the host organism.

The in vivo methods may also be performed in a manner, wherein steps a) and b) are performed sequentially. Thus, the methods may comprise the steps of

I. Providing a host organism comprising

- a. A heterologous nucleic acid encoding a diTPS of class II, which may be any of diTPS of class II described herein in any of the sections "diTPS of class II", "syn-CPP type diTPS", "ent-CPP type diTPS", "(+)-CPP type diTPS", "LPP type diTPS", and "LPP like type diTPS",

II. Incubating said host organism in the presence of GGPP under conditions allowing growth of said host organism, thereby producing a diterpene pyrophosphate intermediate

III. Providing a host organism comprising

- a. A heterologous nucleic acid encoding a diTPS of class I, which may be any of diTPS of class I described herein in any of the sections "diTPS of class I", "EpTPS8", "EpTPS23", "SsSCS", "CfTPS3", "CfTPS4", "MvTPS5", "TwTPS2", "EpTPS1", and "CfTPS14"

- IV. Incubating said host organism in the presence of the diterpene pyrophosphate intermediate produced in step II. under conditions allowing growth of said host organism, thereby producing a diterpene
- V. Optionally isolating the diterpene.

5

In preferred embodiments of the invention the host organism is capable of producing GGPP. Thus step II. may simply be performed by cultivating said host organism. Many host organisms produce GGPP endogenously. Thus, the host organism may be a host organism, which endogenously produce GGPP. Such host organisms for example
10 include plants and yeast. Even if the host organism produce GGPP endogenously, the host organism may be recombinantly modulated to upregulate production of GGPP.

It is also comprised within the invention that GGPP is introduced to the host organism. If the host organism is a microorganism, then GGPP may be added to the cultivation
15 medium of said microorganism. If the host organism is a plant, then GGPP may be added to the growing soil of the plant or it may be introduced into the plant by infiltration. Thus, if the heterologous nucleic(s) are introduced into the plant by infiltration, then GGPP may be co-infiltrated together with the heterologous nucleic acid(s).

20

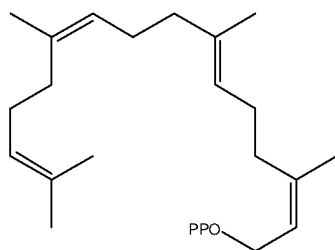
In order to produce a specific diterpene according to the present invention, a useful combination of a diTPS of class II and a diTPS of class I must be employed. Examples of specific combinations of a diTPS of class II and a diTPS of class I, which leads to production of specific diterpenes are shown in figure 2. Other combinations of diTPS of
25 class II and diTPS of class I may be used. In general, the diTPS of class II is selected so that it produces a diterpene pyrophosphate intermediate containing a decalin core having the desired stereochemistry at the 9 and 10 substitutions. Useful diTPS of class II are described below and also specific diTPS of class II catalysing formation of diterpene pyrophosphate intermediates with a specific stereochemistry are described.

30 The diTPS of class I is selected so that it catalyses the conversion of the diterpene pyrophosphate intermediate to the desired diterpene. Useful diTPS of class I are described below. Also specific reactions catalysed by various diTPS of class I are described, enabling the skilled person to select a useful diTPS of class I for production of a desired diterpene. Once a useful diTPS of class II and diTPS of class I have been
35 selected, nucleic acids encoding same may be expressed in the host organism

allowing production of the diterpene in the host organism. Putative useful combinations of a diTPS of class II and a diTPS of class I for production of a given diterpene may be tested by expressing said diTPS of class II and said diTPS of class I in a host organism followed by testing for production of the diterpene, e.g. by GC-MS analysis and/or
5 NMR analysis. Putative useful combinations of a diTPS of class II and a diTPS of class I for production of a given diterpene may in particular be tested as described in Example 1 herein below. Methods for expression of enzymes in host organisms are well known to skilled person, and may for example include the methods described herein below in the section "Heterologous nucleic acids".

10

The term GGPP as used herein refers to geranylgeranyl diphosphate and is a compound of the following structure:



15

wherein PPO- is diphosphate. PPO- and -OPP may be used interchangeably herein.

diTPS of class II

20

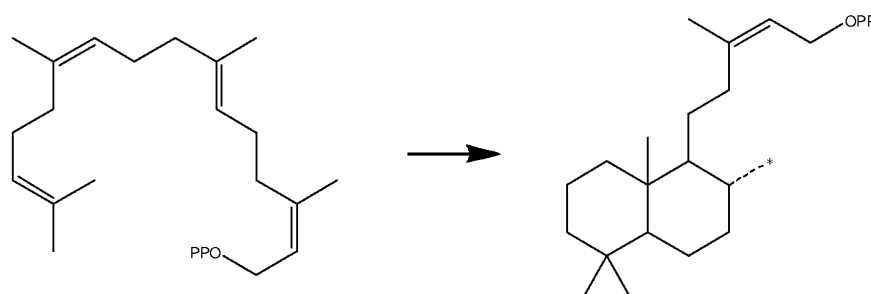
The methods of the invention comprise step a), which involves use of a diTPS of class II. The invention also features host organisms comprising a heterologous nucleic acid encoding a diTPS of class II. The invention also relates to certain diTPS of class II per se.

25

Said diTPS of class II is an enzyme capable of catalysing protonation-initiated cationic cycloisomerization of GGPP to form a diterpene pyrophosphate intermediate. The class II diTPS reaction, may be terminated either by deprotonation or by water capture of the diphosphate carbocation.

30

In particular the diTPS of class II may be an enzyme capable of catalysing the reaction I:



wherein PPO- is diphosphate and the \cdots^* indicates either a double bond or two single bonds, wherein one is substituted with $-\text{OH}$ and the other with $-\text{CH}_3$.

5 Thus, \cdots^* may be CH_2 or $\begin{matrix} \text{CH}_3 \\ \diagup \\ \text{OH} \end{matrix}$.

When no stereochemistry is indicated, the bond may be in any conformation. By selecting appropriate diTPS of class II the stereochemistry of the diterpene produced may be controlled. Accordingly, by following the description of the present invention,
10 the skilled person may be able to design the production of a given diterpene by selecting appropriate diTPS enzymes of class II and class I as described herein.

The diTPS of class II is generally a polypeptide sharing at least some sequence similarity to at least one of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4,
15 SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 or SEQ ID NO:8. In particular, it is preferred that the diTPS of class II shares at least 30%, preferably at least 40% sequence identity with at least one of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 and SEQ ID NO:8. In particular, it is preferred that the diTPS of class II shares at least 30%, such as at least 35%
20 sequence identity to the sequence of SsLPPS (SEQ ID NO:6) or to the sequence of AtCPS (see figure 5). Furthermore, it is preferred that the diTPS of class II in addition to above mentioned sequence identity also contains the following motif of four amino acids:

D/E-X-D-D,

25 wherein X may be any amino acid, such as any naturally occurring amino acids. In particular, X may be an amino acid with a hydrophobic side chain, and thus X may for example be selected from the group consisting of A, I, L, M, F, W, Y and V. Even more preferably X is an amino acid with a small hydrophobic side chain, and thus X may be selected from the group consisting of A, I, L and V.

In one embodiment of the invention said motif of four amino acids is:

D/E-I/V-D-D

5

D/E indicates that said amino acid may be D or E and I/V indicates that said amino acid may be I or V.

Amino acids are herein named using the IUPAC nomenclature for amino acids.

10

In particular, it is preferred that the diTPS of class II contains above described motif in a position corresponding to position aa 372 to 375 of SsLPPS of SEQ ID NO:6. A position corresponding to position aa 372 to 375 of SsLPPS of SEQ ID NO:6 is identified by aligning the sequence of a diTPS of class II of interest to SEQ ID NO:6 and optionally to additional sequences of diTPS of class II as e.g. shown in figure 5 and identifying the amino acids of said diTPS of class II aligning with aa 372 to 375 of SsLPPS of SEQ ID NO:6.

15

It is furthermore preferred that in addition to sharing above mentioned sequence identity and containing said motif, then as many as possible of the amino acids marked with a black box in figure 5 are retained. Thus, when aligned to the sequence of ScLPPS (SEQ ID NO:6), then preferably the diTPS of class II also contains at least 80%, more preferably at least 90%, for example at least 95%, such as all of the amino acids marked by a black box in figure 5. Alternatively, when aligned to the sequence of sequence of AtCPS (see figure 5), then preferably the diTPS of class II also contains at least 80%, more preferably at least 90%, for example at least 95%, such as all of the amino acids marked by a black box in figure 5.

20

25

Thus, the diTPS of class II may for example be selected from the group consisting of diTPS of class II of the following types:

30

- i. syn-CPP type, such as any of the enzymes described herein below in the section "syn-CPP type diTPS"
- ii. ent-CPP type, such as any of the enzymes described herein below in the section "ent-CPP type diTPS"

- iii. (+)-CPP type, such as any of the enzymes described herein below in the section "(+)-CPP type diTPS"
- iv. LPP type, such as any of the such as any of the enzymes described herein below in the section "LPP type diTPS"
- 5 v. LPP like type, such as any of the enzymes described herein below in the section "LPP like type diTPS"

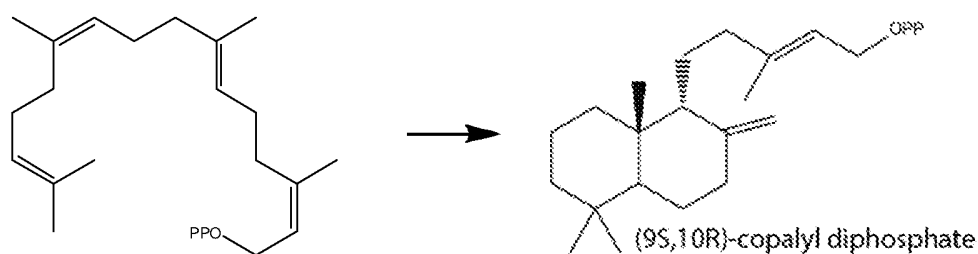
Certain diTPS enzymes are bifunctional in the sense that they may be classified as both class II and class I diTPS enzymes. Such bifunctional diTPS enzymes in general contain both the four amino acids motif: D/E-X-D-D, described herein above, as well as the five amino acid motif: D-D-X-X-D/E, described herein below. It is preferred that the diTPS of class II is not a bifunctional enzyme of both class II and class I. It is also preferred that the diTPS of class I is not a bifunctional enzyme of both class II and class I.

15

syn-CPP type diTPS

The methods of the invention comprise step a), which involves use of a diTPS of class II. The invention also features host organisms comprising a heterologous nucleic acid encoding a diTPS of class II. The invention also relates to certain diTPS of class II per se. In one embodiment said diTPS of class II is a syn-CPP type diTPS. Such diTPS of class II are in particular useful in embodiments of the inventions, wherein the diterpene to be produced contains a 9S,10R decalin core.

25 As used herein the term "syn-CPP type diTPS" refers to any enzyme capable of catalysing the reaction II:



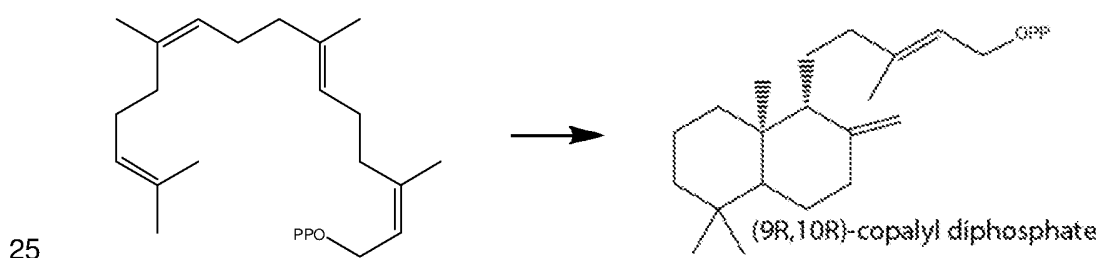
30 wherein PPO- refers to diphosphate.

In one embodiment the syn-CPP type diTPS may be syn-copalyl pyrophosphate synthase (syn-CPP), such as syn-CPP from *Oryza sativa*. In particular, said syn-CPP type diTPS may be a polypeptide of SEQ ID NO:1 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith. The sequence identity is preferably calculated as described herein below in the section "Sequence identity". A functional homologue of a syn-CPP is a polypeptide, which is also capable of catalysing reaction II described above.

ent-CPP type

The methods of the invention comprise step a), which involves use of a diTPS of class II. The invention also features host organisms comprising a heterologous nucleic acid encoding a diTPS of class II. The invention also relates to certain diTPS of class II per se. In one embodiment said diTPS of class II is an ent-CPP type diTPS. Such diTPS of class II are in particular useful in embodiments of the inventions, wherein the diterpene to be produced contains a 9R,10R decalin core.

As used herein the term "ent-CPP type diTPS" refers to any enzyme capable of catalysing the reaction III:



wherein PPO- refers to diphosphate.

In one embodiment the ent-CPP type diTPS may be EpTPS7. In particular, said ent-CPP type diTPS may be a polypeptide of SEQ ID NO:2 or a functional homologue

thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith.

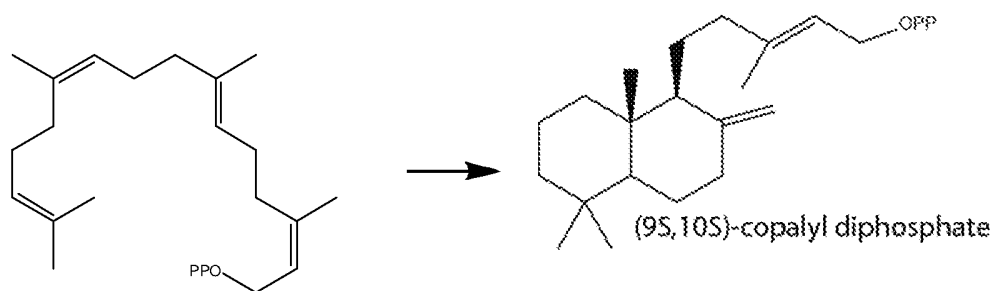
In another embodiment the ent-CPP type diTPS may be ZmAN2. In particular, said ent-CPP type diTPS may be a polypeptide of SEQ ID NO:3 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith.

The sequence identity is preferably calculated as described herein below in the section "Sequence identity". A functional homologue of an ent-CPP is a polypeptide, which is also capable of catalysing reaction III described above.

(+)-CPP type diTPS

The methods of the invention comprise step a), which involves use of a diTPS of class II. The invention also features host organisms comprising a heterologous nucleic acid encoding a diTPS of class II. The invention also relates to certain diTPS of class II per se. In one embodiment said diTPS of class II is a (+)-CPP type diTPS. Such diTPS of class II are in particular useful in embodiments of the inventions, wherein the diterpene to be produced contains a 9S,10S decalin core.

As used herein the term "(+)-CPP type diTPS" refers to any enzyme capable of catalysing the reaction IV:



wherein PPO- refers to diphosphate.

- 5 In one embodiment the (+)-CPP type diTPS may be TwTPS7. In particular, said (+)-CPP type diTPS may be a polypeptide of SEQ ID NO:4 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%,
 10 such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith.

- In another embodiment the (+)-CPP type diTPS may be CfTPS1. In particular, said (+)-CPP type diTPS may be a polypeptide of SEQ ID NO:5 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%,
 15 such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith.

20

The sequence identity is preferably calculated as described herein below in the section "Sequence identity". A functional homologue of a (+)-CPP is a polypeptide, which is also capable of catalysing reaction IV described above.

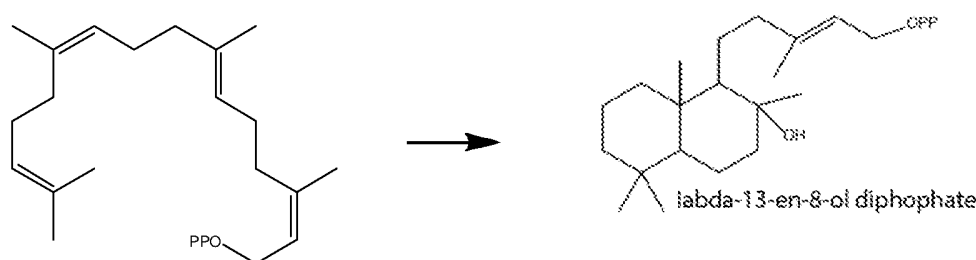
25 **LPP type diTPS**

The methods of the invention comprise step a), which involves use of a diTPS of class II. The invention also features host organisms comprising a heterologous nucleic acid encoding a diTPS of class II. The invention also relates to certain diTPS of class II per

se. In one embodiment said diTPS of class II is a LPP type diTPS. Such diTPS of class II are in particular useful in embodiments of the inventions, wherein the diterpene to be produced contains a 8-hydroxy-decalin core. However, LPP type diTPS may also be useful in other embodiments of the invention.

5

As used herein the term "LPP type diTPS" refers to any enzyme capable of catalysing the reaction V:



10

wherein PPO- refers to diphosphate.

In one embodiment the LPP type diTPS may be labda-13-en-8-ol pyrophosphate synthase, such as SsLPPS. In particular, said LPP type diTPS may be a polypeptide of SEQ ID NO:6 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith. In

embodiments of the invention, wherein the diTPS of class II is SsLPPS or a functional homologue thereof sharing above mentioned sequence identity, then it is preferred that the diTPS of class I is not SsSCS [SEQ ID NO:11], CfTPS3 [SEQ ID NO:12], CfTPS4 [SEQ ID NO:13] or EpTPS8 [SEQ ID NO:9] or a functional homologue of any of the aforementioned sharing at least 70% sequence identity therewith. Thus, in

embodiments of the invention, wherein the diTPS of class II is SsLPPS, then it is preferred that the diTPS of class I is not SsSCS, CfTPS3, CfTPS4 or EpTPS8. It is also preferred that if the diTPS of class II is SsCPSL, then it is preferred that the diTPS of class I is not SsKSL1 or SsKSL2.

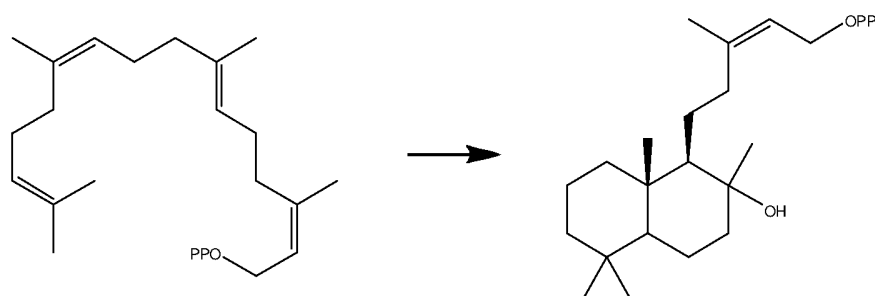
In another embodiment the LPP type diTPS may be TwTPS21. In particular, said LPP type diTPS may be a polypeptide of SEQ ID NO:7 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith.

In another embodiment the LPP type diTPS may be CfTPS2. In particular, said LPP type diTPS may be a polypeptide of SEQ ID NO:17 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith. In embodiments of the invention, wherein the diTPS of class II is CfTPS2 or a functional homologue thereof sharing above mentioned sequence identity, then it is preferred that the diTPS of class I is not CfTPS3 [SEQ ID NO:12] or CfTPS4 [SEQ ID NO:13] or EpTPS8 [SEQ ID NO:9] or a functional homologue of any of the aforementioned sharing at least 70% sequence identity therewith. Thus, in embodiments of the invention, wherein the diTPS of class II is CfTPS2, then it is preferred that the diTPS of class I is not CfTPS3 or CfTPS4 or EpTPS8.

The sequence identity is preferably calculated as described herein below in the section "Sequence identity". A functional homologue of a LPP is a polypeptide, which is also capable of catalysing reaction V described above.

The LLP type diTPS may be an (+)-LPP type diTPS or an ent-LPP type diTPS. Thus, in one embodiment of the invention, the diTPS of class II is an (+)-LPP type diTPS.

As used herein the term "(+)-LPP type diTPS" refers to any enzyme capable of catalysing the reaction XXXIII:

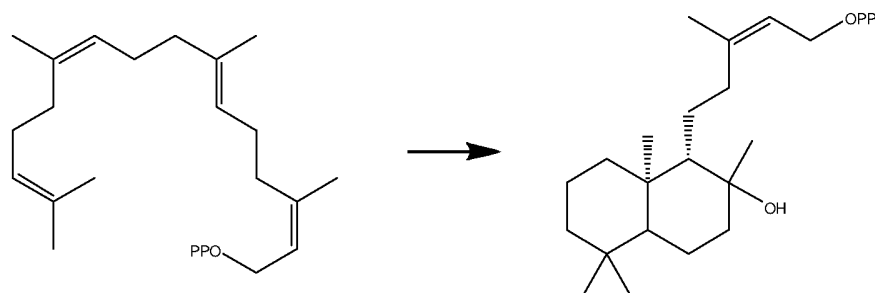


wherein -OPP refers to diphosphate.

In one embodiment the (+)-LPP type diTPS may be labda-13-en-8-ol pyrophosphate synthase, such as SsLPPS. In particular, said (+)-LPP type diTPS may be a polypeptide of SEQ ID NO:6 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith. In embodiments of the invention, wherein the diTPS of class II is SsLPPS or a functional homologue thereof sharing above mentioned sequence identity, then it is preferred that the diTPS of class I is not SsSCS [SEQ ID NO:11], CfTPS3 [SEQ ID NO:12], CfTPS4 [SEQ ID NO:13] or EpTPS8 [SEQ ID NO:9] or a functional homologue of any of the aforementioned sharing at least 70% sequence identity therewith. Thus, in embodiments of the invention, wherein the diTPS of class II is SsLPPS, then it is preferred that the diTPS of class I is not SsSCS, CfTPS3, CfTPS4 or EpTPS8

In one embodiment of the invention, the diTPS of class II is an ent-LPP type diTPS.

As used herein the term "ent-LPP type diTPS" refers to any enzyme capable of catalysing the reaction XXXIV:



wherein -OPP refers to diphosphate.

In one embodiment the ent-LPP type diTPS may be TwTPS21. In particular, said net-LPP type diTPS may be a polypeptide of SEQ ID NO:7 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith.

10

LPP like type diTPS

The methods of the invention comprise step a), which involves use of a diTPS of class II. The invention also features host organisms comprising a heterologous nucleic acid encoding a diTPS of class II. The invention also relates to certain diTPS of class II per se. In one embodiment said diTPS of class II is a LPP like type diTPS.

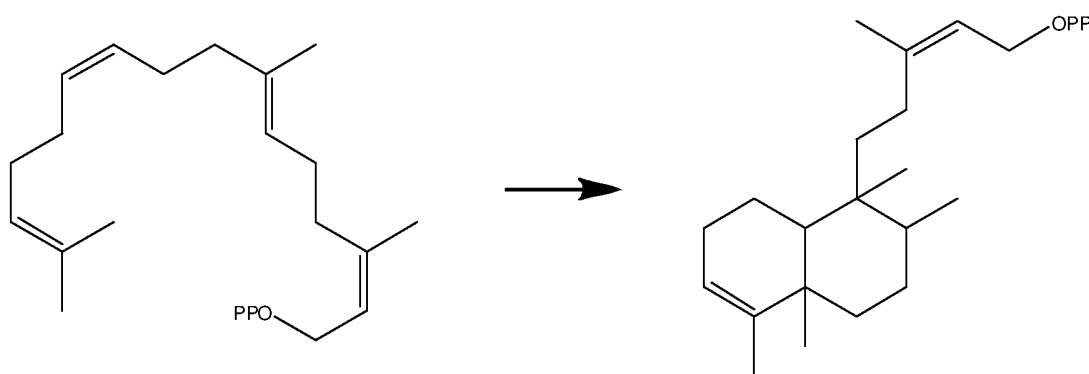
In one embodiment the LPP like type diTPS may be TwTPS14/28. In particular, said LPP like type diTPS may be a polypeptide of SEQ ID NO:8 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith.

25

The LPP like type diTPS may in one embodiment be a CLPP type diTPS.

As used herein the term "CLPP type diTPS" refers to any enzyme capable of catalysing the reaction XXXV:

30

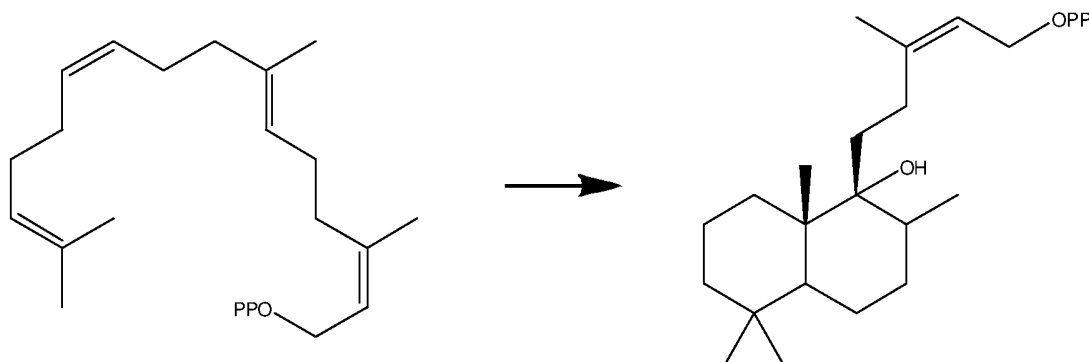


wherein PPO- refers to diphosphate.

- 5 The ClPP type diTPS may for example be TwTPS14/28. In particular, said ClPP type diTPS may be a polypeptide of SEQ ID NO:8 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith. A functional homologue of TwTPS14/28 may in particular be a polypeptide have aforementioned sequence identity with TwTPS14/28 and which also is capable of catalysing reaction XXXV.

- 15 The LPP like type diTPS may in one embodiment be a 9-LPP type diTPS.

As used herein the term "9-LPP type diTPS" refers to any enzyme capable of catalysing the reaction XXXVI:



20

wherein PPO- refers to diphosphate.

The 9-LPP type diTPS may for example be MvTPS1. In particular, said 9-LPP type diTPS may be a polypeptide of SEQ ID NO:28 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith. A functional homologue of MvTPS1 may in particular be a polypeptide have aforementioned sequence identity with MvTPS1 and which also is capable of catalysing reaction XXXVI.

The sequence identity is preferably calculated as described herein below in the section "Sequence identity".

diTPS of class I

The methods of the invention comprise step b), which involves use of a diTPS of class I. The invention also features host organisms comprising a heterologous nucleic acid encoding a diTPS of class I. The invention also relates to certain diTPS of class I per se.

Said diTPS of class I is an enzyme capable of catalyzing cleavage of the diphosphate group of the diterpene pyrophosphate intermediate and additionally preferably also is capable of catalysing cyclization and/or rearrangement reactions on the resulting carbocation. As with the class II diTPSs, deprotonation or water capture may terminate the class I diTPS reaction leading to hydroxylation of the diterpene pyrophosphate intermediate.

The diTPS of class I is generally a polypeptide sharing at least some sequence similarity to at least one of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16 or SEQ ID NO:17. In particular, it is preferred that the diTPS of class I shares at least 30%, preferably at least 40%, more preferably at least 45% sequence identity with at least one of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16 and SEQ ID NO:17. In particular, it is preferred that the diTPS of class I shares at least 30%, such as at least 35% sequence identity to the sequence of ScSCS (SEQ ID NO:11) or to the sequence

of AtEKS (see figure 4). Furthermore, it is preferred that the diTPS of class I in addition to above mentioned sequence identity also contains the following motif of five amino acids:

D-D-X-X-D/E,

5 wherein X may be any amino acid, such as any naturally occurring amino acids. In particular, X may be an amino acid with a hydrophobic side chain, and thus X may for example be selected from the group consisting of A, I, L, M, F, W, Y and V. Even more preferably X is an amino acid with a small hydrophobic side chain, and thus X may be selected from the group consisting of A, I, L and V.

10

In one embodiment of the invention said motif of five amino acids is:

D-D-F-F-D/E

D/E indicates that said amino acid may be D or E.

15

In particular, it is preferred that the diTPS of class I contains said motif in a position corresponding to position aa 329-333 of SsSCS of SEQ ID NO:11. A position corresponding to position aa 329-333 of SsSCS of SEQ ID NO:11 is identified by aligning the sequence of a diTPS of class I of interest to SEQ ID NO:11 and optionally
20 to additional sequences of diTPS of class I as e.g. shown in figure 4, and identifying the amino acids of said diTPS of class I aligned with aa 329-333 of SsSCS of SEQ ID NO:11.

It is furthermore preferred that in addition to sharing above mentioned sequence
25 identity and containing said motif, then as many as possible of the amino acids marked with a black box in figure 4 are retained. Thus, when aligned to the sequence of ScSCS (SEQ ID NO:11), then preferably the diTPS of class I also contains at least 80%, more preferably at least 90%, for example at least 95%, such as all of the amino acids marked by a black box in figure 4. Alternatively, when aligned to the sequence of
30 sequence of AtEKS (see figure 4), then preferably the diTPS of class I also contains at least 80%, more preferably at least 90%, for example at least 95%, such as all of the amino acids marked by a black box in figure 4.

Thus, the diTPS of class I may for example be selected from the group consisting of
35 diTPS of class I of the following types:

- i. EpTPS8 like diTPS, such as any of the enzymes described herein below in the section "EpTPS8"
- ii. EpTPS23 like diTPS, such as any of the enzymes described herein below in the section "EpTPS23"
- 5 iii. SsSCS like diTPS, such as any of the enzymes described herein below in the section "SsSCS"
- iv. CfTPS3 like diTPS, such as any of the enzymes described herein below in the section "CfTPS3"
- v. CfTPS4 like diTPS, such as any of the enzymes described herein below in the section "CfTPS4"
- 10 vi. TwTPS2 like diTPS, such as any of the enzymes described herein below in the section "TwTPS2"
- vii. EpTPS1 like diTPS, such as any of the enzymes described herein below in the section "TwTPS1"
- 15 viii. CfTPS14 like diTPS, such as any of the enzymes described herein below in the section "CfTPS14"

The diTPS of class I may in one embodiment also be MvTPS5 like diTPS, such as any of the enzymes described herein below in the section "MvTPS5".

20

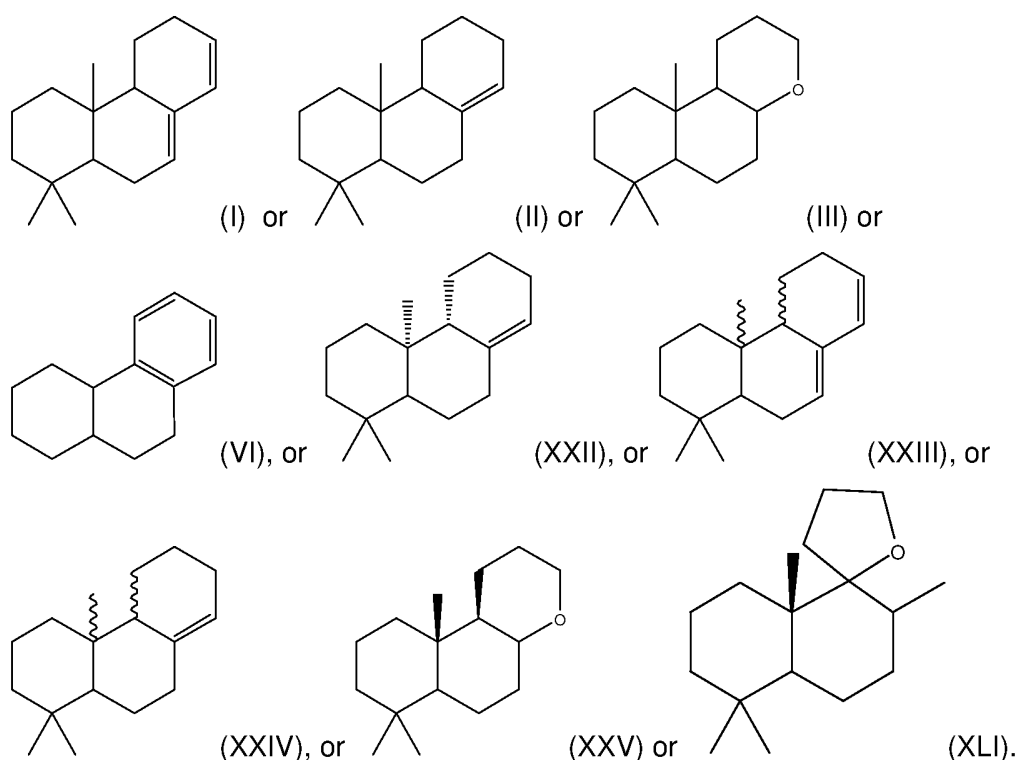
EpTPS8

The invention involves use of a diTPS of class I. In one embodiment said diTPS of class I may be an EpTPS8 like diTPS. In embodiments of the invention, wherein the diTPS of class I is a EpTPS8 like diTPS, then it is preferred that the diTPS of class II is not CfTPS2[SEQ ID NO:17], or SsLPPS [SEQ ID NO:6] or a functional homologue of any of the aforementioned sharing at least 70% sequence identity therewith. Thus, in embodiments of the invention, wherein the diTPS of class I is EpTPS8, then it is preferred that the diTPS of class II is not CfTPS2 or SsLPPS.

30

In particular, said diTPS of class I may be an EpTPS8 like diTPS in embodiments of the invention, wherein the diterpene to be produced contains a tricyclic ring structure. For example said diTPS of class I may be and EpTPS8 like diTPS in embodiments of the invention, wherein the diterpene to be produced contains a core of any of the formulas I, II, III, VI, XXII, XXIII, XXIV or XXV:

35



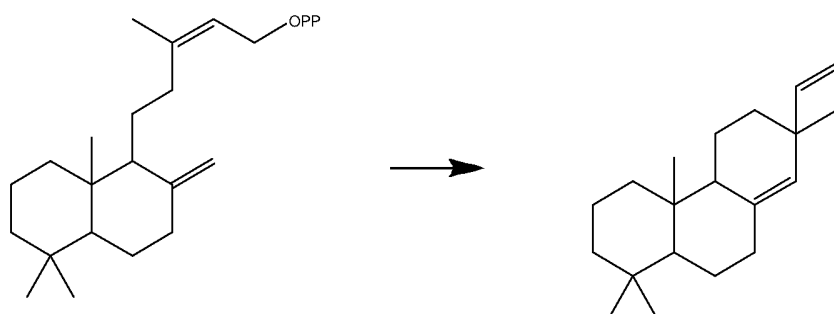
- 5 The wavy line “~” as used herein indicates a bond of undefined stereochemistry, i.e. the bond may be either a “!” or “~”.

Dependent on the structure of the diterpene pyrophosphate intermediate then the diterpene containing a core of formula I or II may have different stereochemistry. In
 10 general the stereochemistry of the decalin core present in the diterpene pyrophosphate intermediate is maintained after the reaction catalysed by a EpTPS8 like diTPS.

The EpTPS8 like diTPS may be any enzyme capable of catalysing the reaction VII:

- 15 Diterpene pyrophosphate intermediate containing a decalin core structure \longrightarrow
 Diterpene containing a core structure of formula I or formula II or formula III or formula VI.

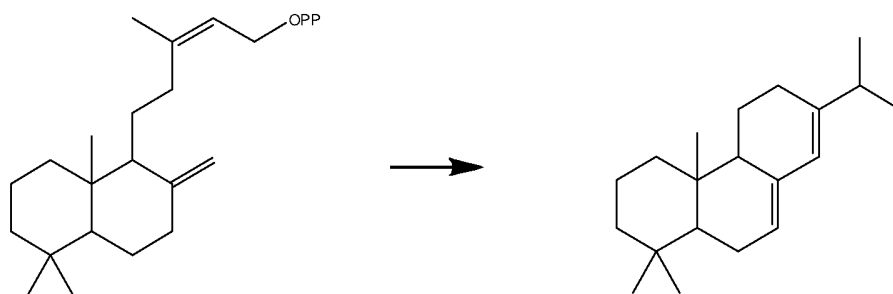
In particular EpTPS8 like diTPS may be an enzyme catalysing the reaction VIII:



wherein -OPP indicates diphosphate. During reaction VIII the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

5

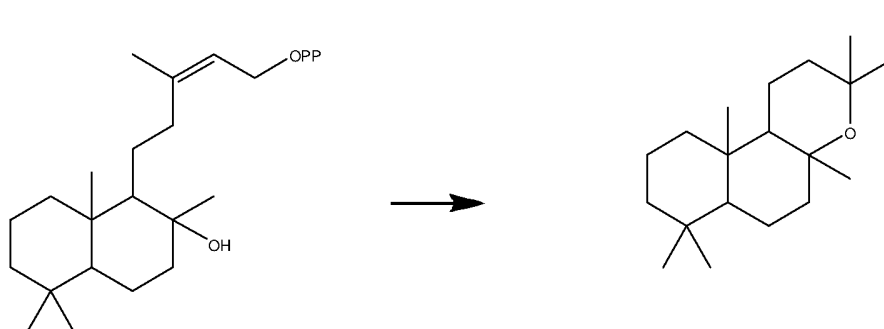
The EpTPS8 like diTPS may also be an enzyme catalysing the reaction IX:



10

wherein OPP indicated diphosphate. During reaction IX the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

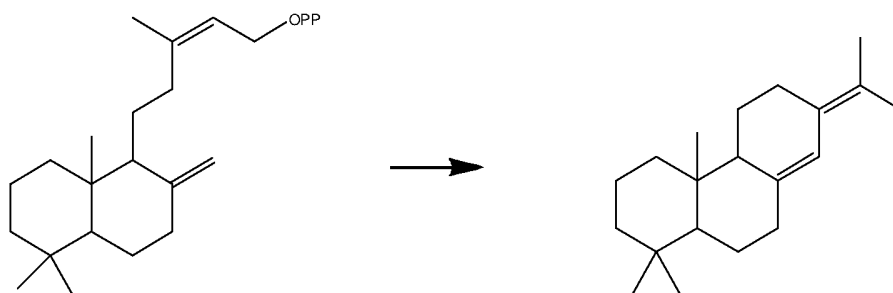
The EpTPS8 like diTPS may also be an enzyme catalysing the reaction X:



15

wherein -OPP indicated diphosphate. During reaction X the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

In particular, the EpTPS8 like diTPS may be an enzyme catalysing the reaction XXV:



5 wherein -OPP indicates diphosphate. During reaction XXV the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

10 In one embodiment EpTPS8 like diTPS may be a terpene synthase from *Euphobia peplus*, and in particular it may be TPS8 from *Euphobia peplus*. TPS8 from *Euphobia peplus* is also referred to as EpTPS herein. In particular, said EpTPS8 like diTPS may be a polypeptide of SEQ ID NO:9 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as
15 at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith.

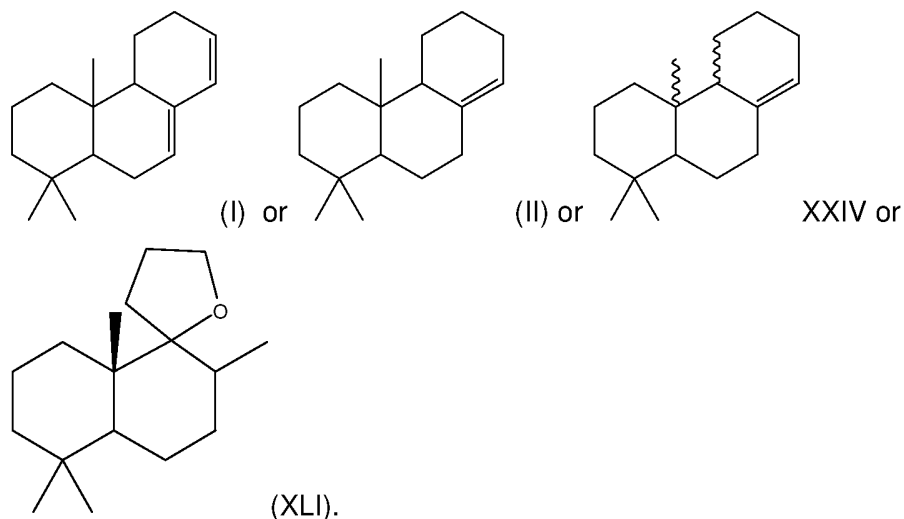
20 The sequence identity is preferably calculated as described herein below in the section "Sequence identity". A functional homologue of EpTPS8 is a polypeptide, which is also capable of catalysing at least one of reactions VII, VIII, IX, X and XXV described above.

EpTPS23

25 The invention involves use of a diTPS of class I. In one embodiment said diTPS of class I may be an EpTPS23 like diTPS.

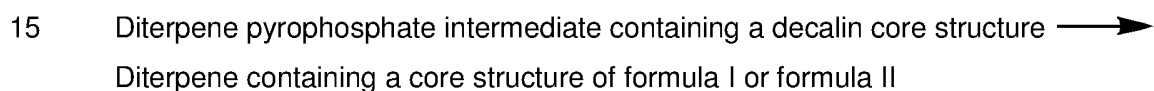
In particular, said diTPS of class I may be an EpTPS23 like diTPS in embodiments of the invention, wherein the diterpene to be produced contains a tricyclic ring structure.
30 For example said diTPS of class I may be an EpTPS23 like diTPS in embodiments of

the invention, wherein the diterpene to be produced contains a core of any of the formulas I and II:

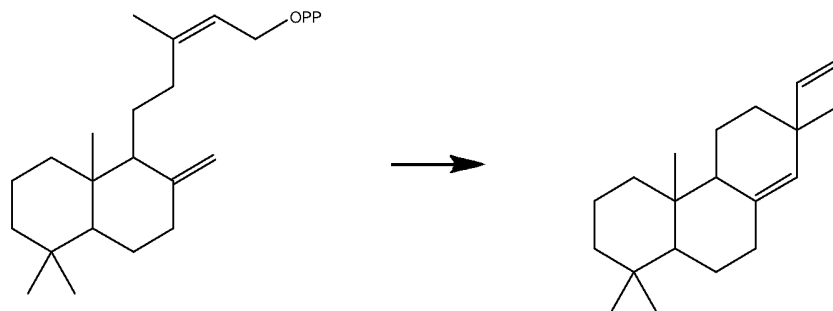


Dependent on the structure of the diterpene pyrophosphate intermediate then the diterpene containing a core of formula I or II may have different stereochemistry. In general the stereochemistry of the decalin core present in the diterpene pyrophosphate intermediate is maintained after the reaction catalysed by an EpTPS23 like diTPS.

The EpTPS23 like diTPS may in particular be an enzyme capable of catalysing the reaction XI:

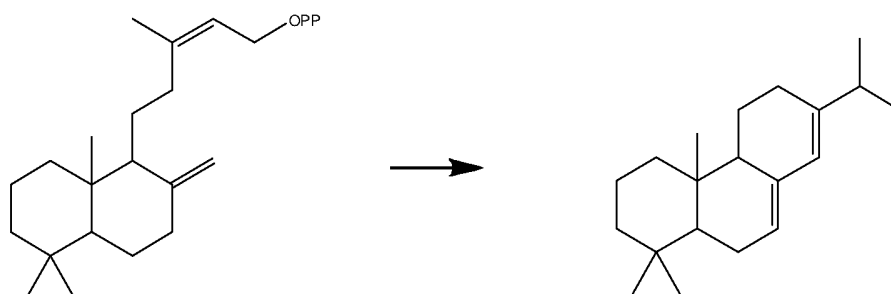


In particular an EpTPS23 like diTPS may be an enzyme catalysing the reaction VIII:



wherein -OPP indicated diphosphate. During reaction VIII the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

- 5 The EpTPS23 like diTPS may also be an enzyme catalysing the reaction IX:



- 10 wherein -OPP indicated diphosphate. During reaction IX the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

- In one embodiment an EpTPS23 like diTPS may be a diterpene synthase from *Euphobia peplus*. In particular, the EpTPS23 like diTPS may be TPS23 of *Euphobia*
15 *peplus*. TPS23 of *Euphobia peplus* may also be referred to as EpTPS23 herein. In particular, said EpTPS23 like diTPS may be a polypeptide of SEQ ID NO:10 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%,
20 such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith.

- The sequence identity is preferably calculated as described herein below in the section "Sequence identity". A functional homologue of EpTPS23 is a polypeptide, which is
25 also capable of catalysing at least one of reactions VIII or IX described above.

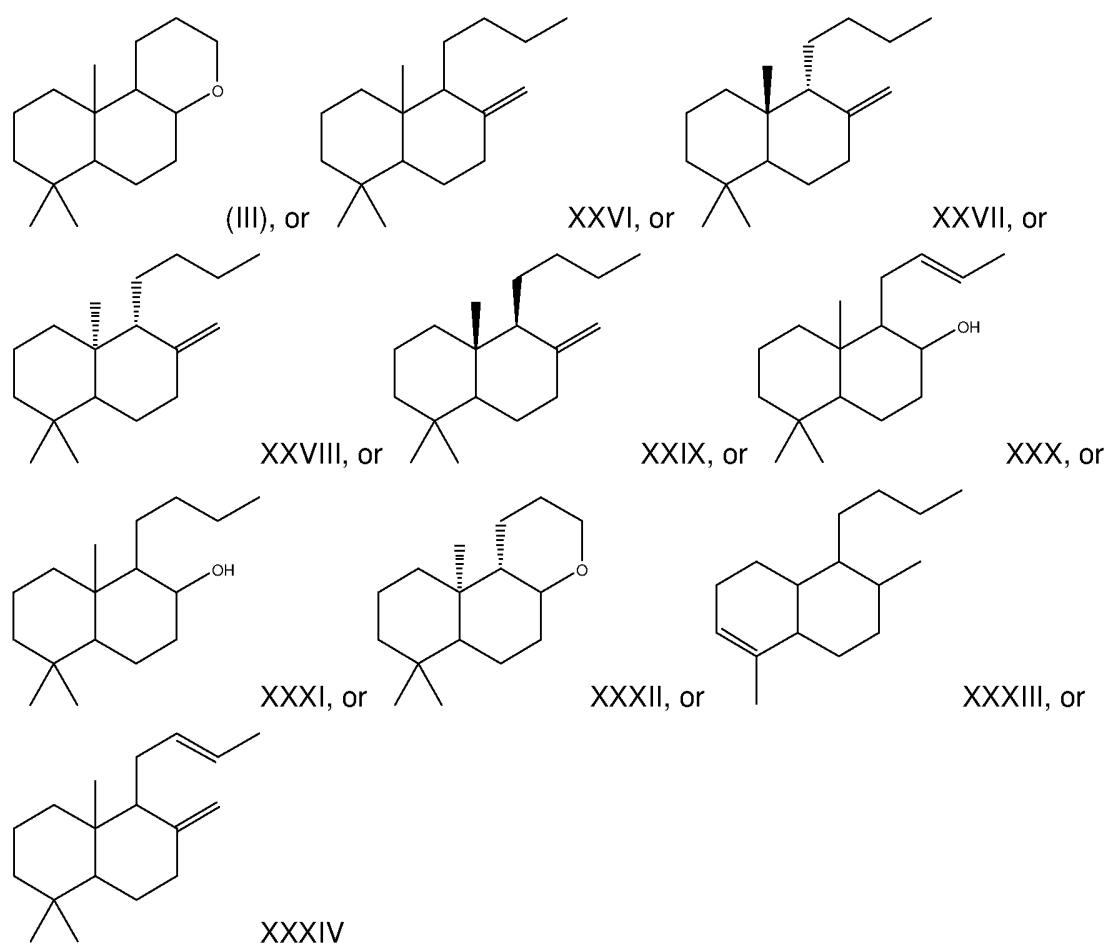
SsSCS

- The invention involves use of a diTPS of class I. In one embodiment said diTPS of
30 class I may be a SsSCS like diTPS.

In particular, said diTPS of class I may be a SsSCS like diTPS in embodiments of the invention, wherein the diterpene to be produced contains a decalin substituted at the 10 position with C₅-alkenyl chain, which optionally may be substituted with a hydroxyl
 5 and/or a methyl group and/or =C.

Furthermore, said diTPS of class I may be a SsSCS like diTPS in embodiments of the invention, wherein the diterpene to be produced contains a core of formula III, XXVI, XXVII, XXVIII, XXIX, XXX, XXXI, XXXII, XXXIII, or XXXIV:

10



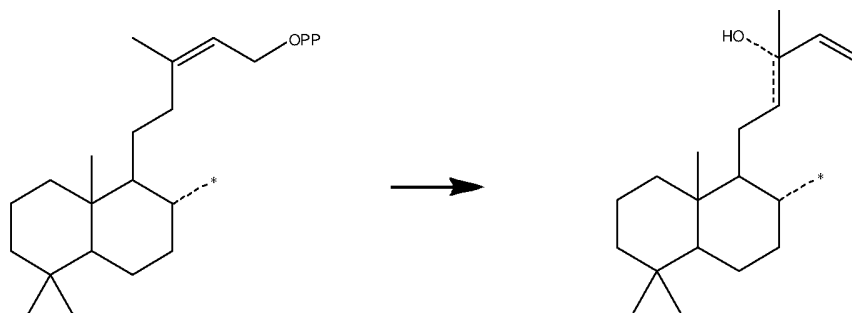
15

Dependent on the structure of the diterpene pyrophosphate intermediate then the diterpene containing a decalin substituted at the 10 position with said C₅-alkenyl chain, or the diterpene containing a core of formula III may have different stereochemistry. In general the stereochemistry of the decalin core present in the diterpene pyrophosphate
 20 intermediate is maintained after the reaction catalysed by a SsSCS like diTPS.

The SsSCS like diTPS may be any enzyme capable of catalysing the following reaction XII:

- 5 Diterpene pyrophosphate intermediate containing a decalin core structure \longrightarrow
 Diterpene containing a decalin core substituted at the 10 position with C₅-alkenyl chain,
 which optionally may be substituted with a hydroxyl and/or a methyl group and/or =C
 OR diterpene containing a core structure of formula III.

- 10 The SsSCS like diTPS may in particular be an enzyme capable of catalysing the
 reaction XVI:



wherein -OPP is diphosphate; and

.....* indicates either a double bond or two single bonds, wherein one is substituted with

- 15 -OH and the other with -CH₃; and

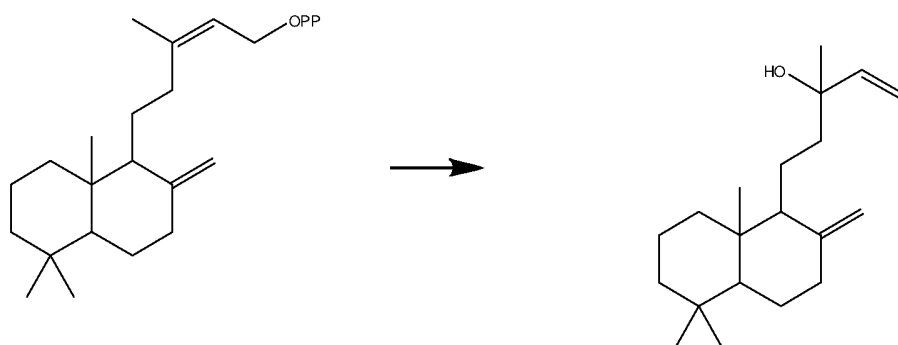
the dotted lines without star indicates a bond, which optionally is present.

Thus,* may be =CH_2 or CH_3 or OH .

- 20 It is to be understood that in embodiments of the invention, wherein the dotted line

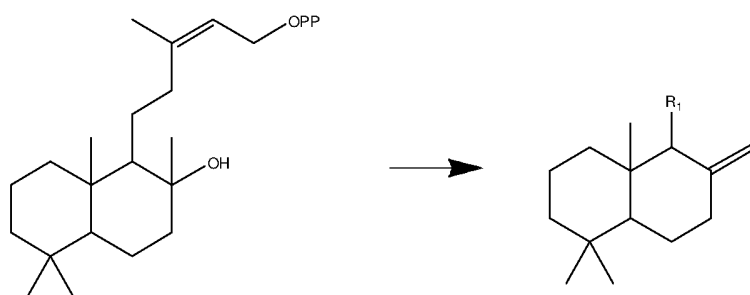
shown as OH is not present, then also the hydroxyl group is not present. It is preferred
 that one and only one of the dotted lines without star indicates a bond.

- 25 A SsSCS like diTPS may in particular be an enzyme capable of catalysing the reaction
 XVII:



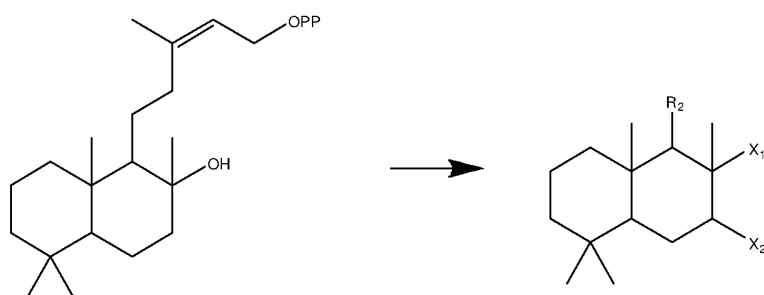
wherein OPP indicated diphosphate. During reaction XVII the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate. Thus, the SsSCS like diTPS may be an enzyme catalysing any of the reactions XIII, XIV and XV shown in figure 1.

The SsSCS like diTPS may also be an enzyme catalysing the following reaction XXVIII:



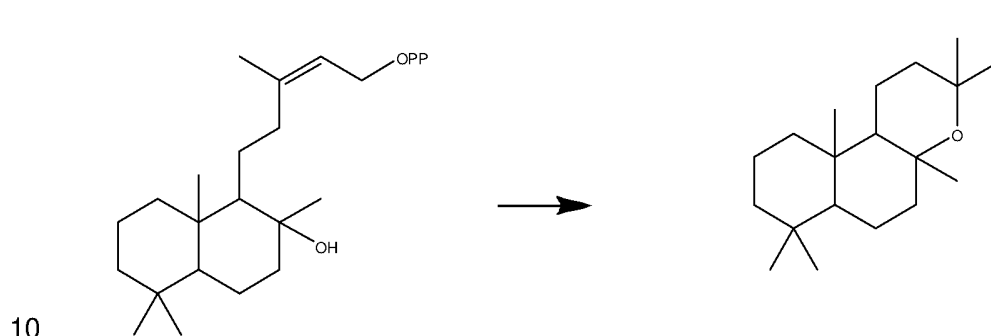
wherein OPP is diphosphate and R_1 is a C_5 -alkenyl substituted with methyl and/or hydroxyl. Preferably, R_1 is C_5 -alkenyl containing one or two double bonds. When R_1 is alkenyl containing one double bond, said alkenyl is preferably substituted with hydroxyl and methyl. When R_1 is alkenyl containing two double bonds, said alkenyl is preferably substituted with methyl.

The SsSCS like diTPS may also be an enzyme catalysing the following reaction XXIX:



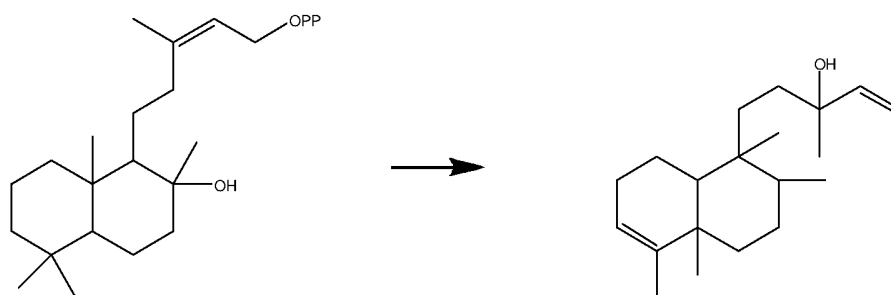
wherein -OPP is diphosphate and R_2 is a C_5 -alkenyl substituted with methyl and/or hydroxyl or with =C, and X_1 is either -OH or methyl, and X_2 is either -H or -OH, wherein one and only one of X_1 and X_2 is -OH. Preferably, R_2 is C_5 -alkenyl containing one or two double bonds. When R_2 is alkenyl containing one double bond, said alkenyl is preferably substituted with hydroxyl and methyl or with =C. When R_2 is alkenyl containing two double bonds, said alkenyl is preferably substituted with methyl.

The SsSCS like diTPS may also be an enzyme catalysing the reaction X:



wherein OPP indicates diphosphate. During reaction X the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

15 The SsSCS like diTPS may also be an enzyme catalysing the reaction XXX:



wherein OPP indicates diphosphate.

20 In one embodiment a SsSCS like diTPS may be Sclareol Synthase (SCS) from *Salvia Sclarea*. SCS from *Salvia Sclarea* may also be referred to as SsSCS herein. In particular, said SsSCS like diTPS may be a polypeptide of SEQ ID NO:11 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such

as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith.

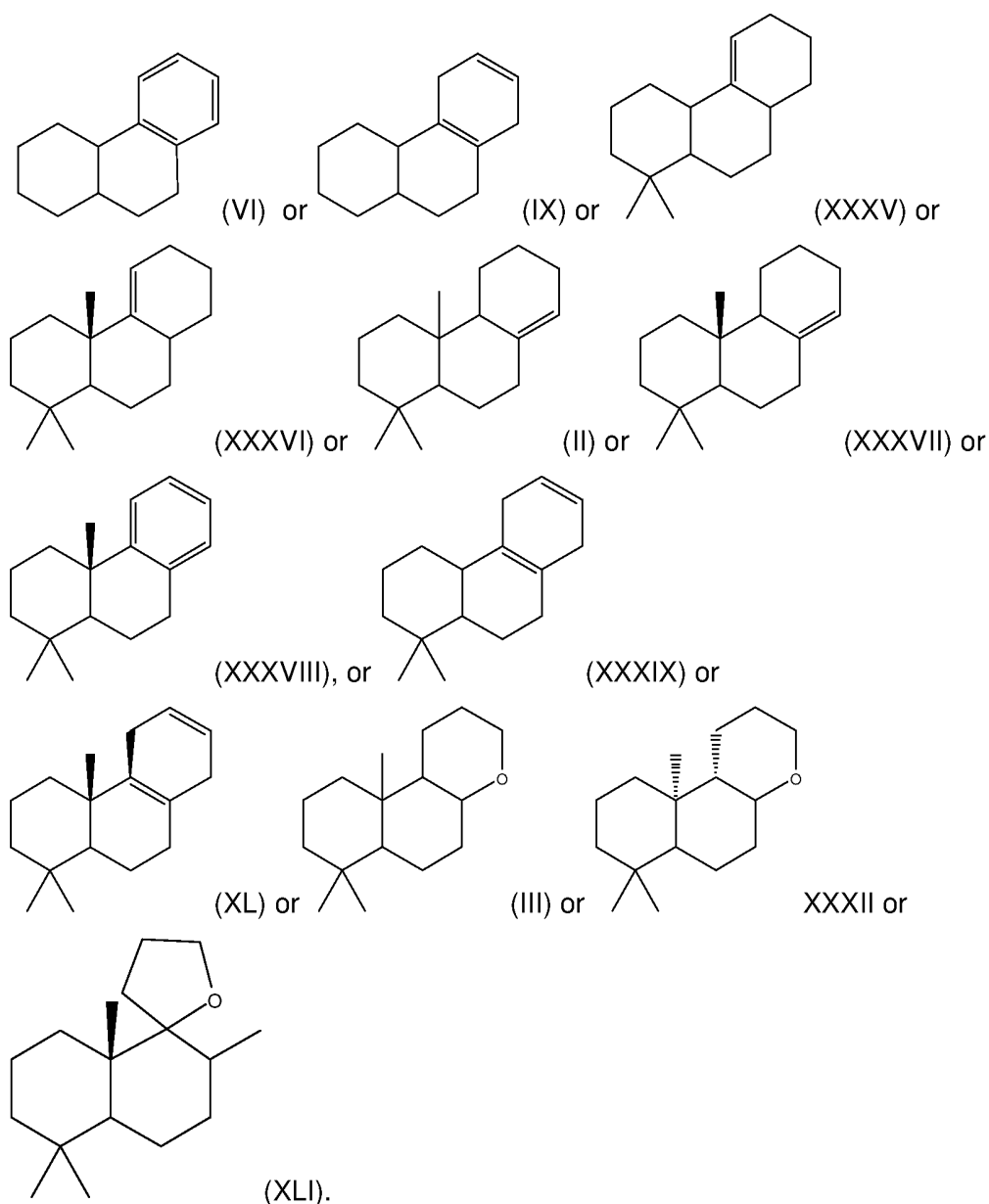
- 5 The sequence identity is preferably calculated as described herein below in the section "Sequence identity". A functional homologue of SsSCS is a polypeptide, which is also capable of catalysing at least one of reactions XII, XIII, XIV, XV, XVI, XVII, XXVIII, XXIX, or XXX described above.

10 **CfTPS3**

The invention involves use of a diTPS of class I. In one embodiment said diTPS of class I may be a CfTPS3 like diTPS. In embodiments of the invention, wherein the diTPS of class I is a CfTPS3 like diTPS, then it is preferred that the diTPS of class II is
15 not CfTPS2 [SEQ ID NO:17], or SsLPPS [SEQ ID NO:6] or a functional homologue of any of the aforementioned sharing at least 70% sequence identity therewith. Thus, in embodiments of the invention, wherein the diTPS of class I is CfTPS3, then it is preferred that the diTPS of class II is not CfTPS2 or SsLPPS.

20 In particular, said diTPS of class I may be a CfTPS3 like diTPS in embodiments of the invention, wherein the diterpene to be produced contains a tricyclic ring structure. For example said diTPS of class I may be a CfTPS3 like diTPS in embodiments of the invention, wherein the diterpene to be produced contains a core of any of the formulas VI, IX, XXXV, XXXVI, II,XXXVII, XXXVIII, XXXIX, XL, III or XXXII:

25



5

Dependent on the structure of the diterpene pyrophosphate intermediate then the diterpene containing a core of formula VI, IX, XXXV, II, or XXXIX may have different stereochemistry. In general the stereochemistry of the decalin core present in the diterpene pyrophosphate intermediate is maintained after the reaction catalysed by the CfTPS3 like diTPS.

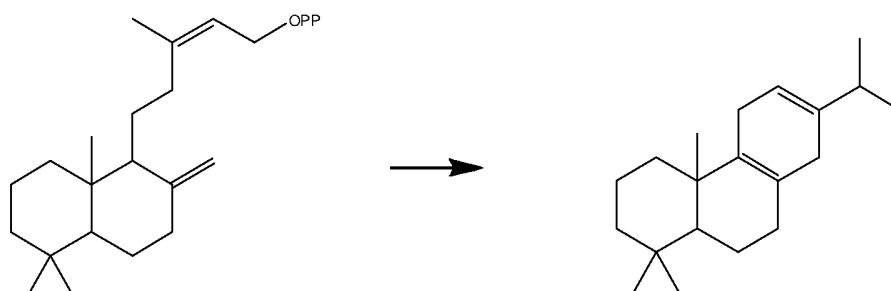
10

The CfTPS3 like diTPS may be any enzyme capable of catalysing the reaction XXIII:

15

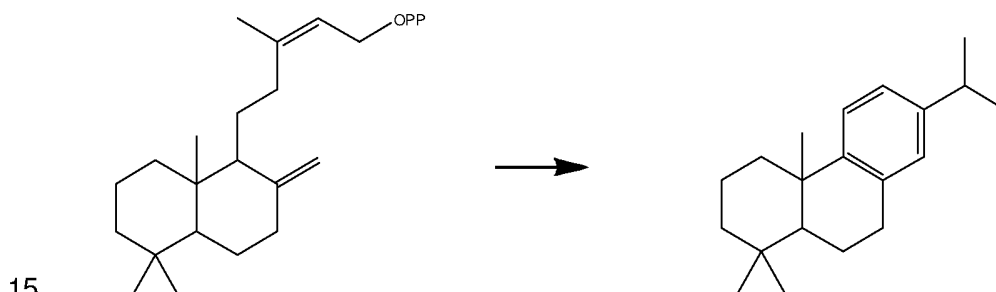
Diterpene pyrophosphate intermediate containing a decalin core structure \longrightarrow
Diterpene containing a core structure of formula VI, formula IX, XXXV, XXXVI,
II, XXXVII, XXXVIII, XXXIX, XL, III or XXXII.

- 5 The CfTPS3 like diTPS may in particular be an enzyme capable of catalysing the reaction XXIV:



wherein OPP indicates diphosphate. During reaction XXIV the produced diterpene will
in general maintain the stereochemistry around the decalin core found in the starting
10 diterpene pyrophosphate intermediate.

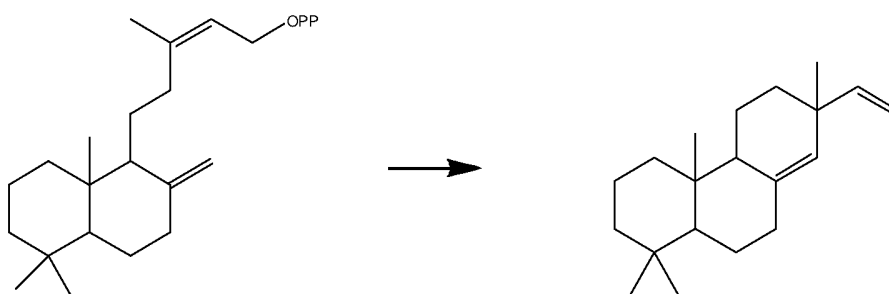
The CfTPS3 like diTPS may in particular be an enzyme capable of catalysing the
reaction XXII:



15 wherein OPP is diphosphate. During reaction XXII the produced diterpene will in
general maintain the stereochemistry around the decalin core found in the starting
diterpene pyrophosphate intermediate.

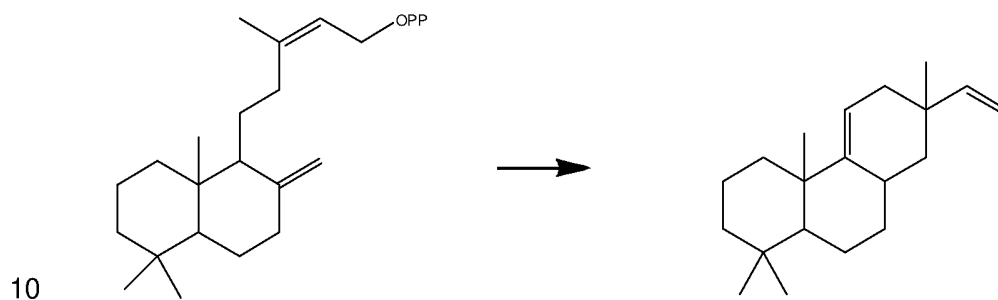
20

The CfTPS3 like diTPS may in particular be an enzyme capable of catalysing the
reaction XXXI:



wherein OPP is diphosphate. During reaction XXXI the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting
5 diterpene pyrophosphate intermediate.

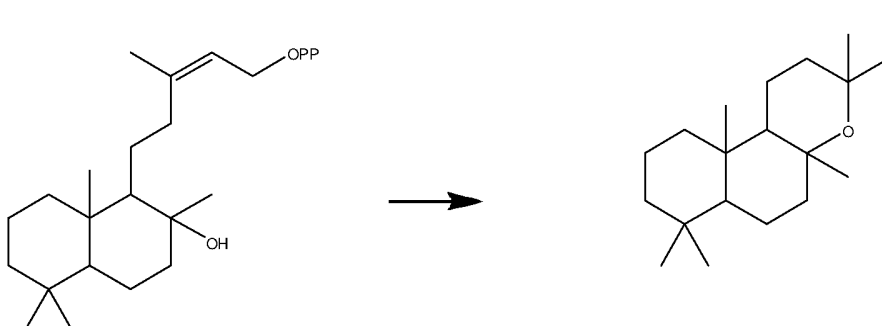
The CfTPS3 like diTPS may in particular be an enzyme capable of catalysing the reaction XXXII:



wherein OPP is diphosphate. During reaction XXXII the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting
10 diterpene pyrophosphate intermediate.

15

The CfTPS3 like diTPS may also be an enzyme catalysing the reaction X:

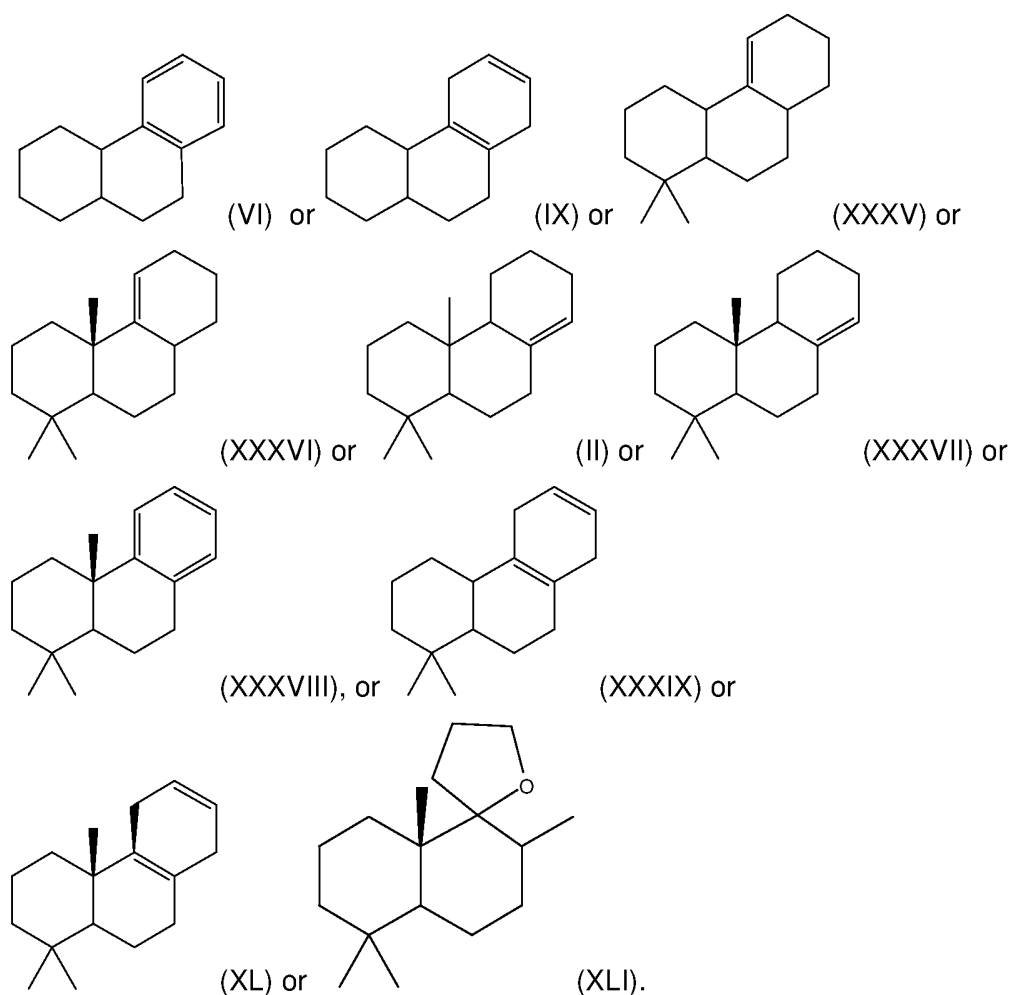


wherein OPP indicates diphosphate. During reaction X the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

- 5 In one embodiment the CfTPS3 like diTPS may be a diterpene synthase from *Coleus forskohlii*. In particular, the CfTPS3 like diTPS may be a TPS3 from *Coleus forskohlii*. TPS3 from *Coleus forskohlii* may also be referred to as CfTPS3. In particular, said CfTPS3 like diTPS may be a polypeptide of SEQ ID NO:12 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as
- 10 at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith.
- 15 The sequence identity is preferably calculated as described herein below in the section "Sequence identity". A functional homologue of CfTPS3 is a polypeptide, which is also capable of catalysing at least one of reactions XXII, XXIII or XXIV described above.

CfTPS4

- 20 The invention involves use of a diTPS of class I. In one embodiment said diTPS of class I may be a CfTPS4 like diTPS. In embodiments of the invention, wherein the diTPS of class I is a CfTPS4 like diTPS, then it is preferred that the diTPS of class II is not CfTPS2[SEQ ID NO:17], or SsLPPS [SEQ ID NO:6] or a functional homologue of any of the aforementioned sharing at least 70% sequence identity therewith. Thus, in
- 25 embodiments of the invention, wherein the diTPS of class I is CfTPS4, then it is preferred that the diTPS of class II is not CfTPS2 or SsLPPS.
- In particular, said diTPS of class I may be a CfTPS4 like diTPS in embodiments of the invention, wherein the diterpene to be produced contains a tricyclic ring structure. For example said diTPS of class I may be a CfTPS4 like diTPS in embodiments of the
- 30 invention, wherein the diterpene to be produced contains a core of any of the formulas VI, IX, XXXV, XXXVI, II, XXXVII, XXXVIII, XXXIX or XL:



5

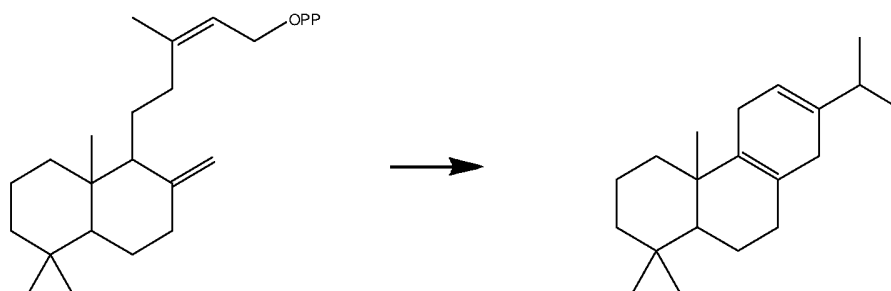
Dependent on the structure of the diterpene pyrophosphate intermediate then the diterpene containing a core of formula VI, IX, XXXV, II, or XXXIX, may have different stereochemistry. In general the stereochemistry of the decalin core present in the diterpene pyrophosphate intermediate is maintained after the reaction catalysed by the CfTPS4 like diTPS.

The CfTPS4 like diTPS may be any enzyme capable of catalysing the reaction XXIII:

15

Diterpene pyrophosphate intermediate containing a decalin core structure \longrightarrow
 Diterpene containing a core structure of formula VI, IX, XXXV, XXXVI, II, XXXVII, XXXVIII, XXXIX or XL.

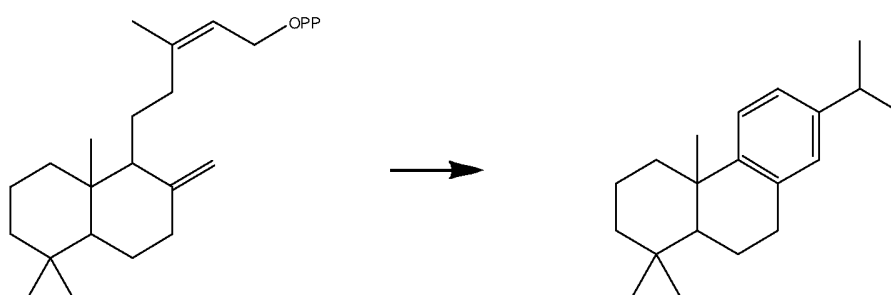
The CfTPS4 like diTPS may in particular be an enzyme capable of catalysing the reaction XXIV:



5 wherein OPP indicates diphosphate. During reaction XXIV the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

The CfTPS4 like diTPS may in particular be an enzyme capable of catalysing the reaction XXII:

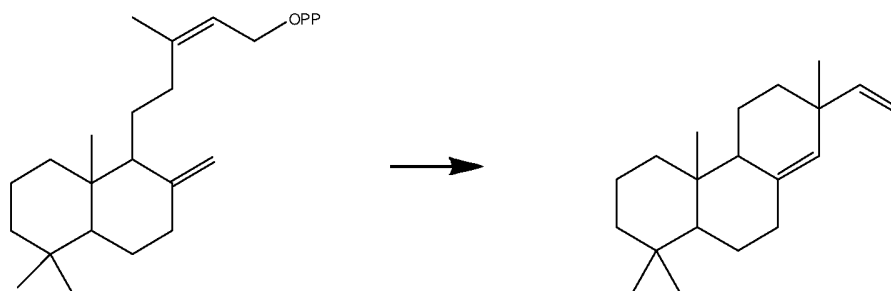
10



15 wherein OPP is diphosphate. During reaction XXII the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

The CfTPS4 like diTPS may in particular be an enzyme capable of catalysing the reaction XXXI:

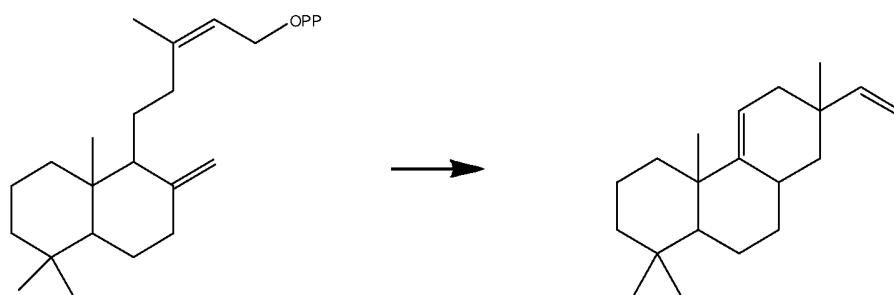
20



wherein OPP is diphosphate. During reaction XXXI the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

5

The CfTPS4 like diTPS may in particular be an enzyme capable of catalysing the reaction XXXII:



10

wherein OPP is diphosphate. During reaction XXXII the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

15 In one embodiment the CfTPS4 like diTPS may be a diterpene synthase from *Coleus forskohlii*. In particular, the CfTPS4 like diTPS may be a TPS4 from *Coleus forskohlii*. TPS4 from *Coleus forskohlii* may also be referred to as CfTPS4. In particular, said CfTPS4 like diTPS may be a polypeptide of SEQ ID NO:13 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as
20 at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith.

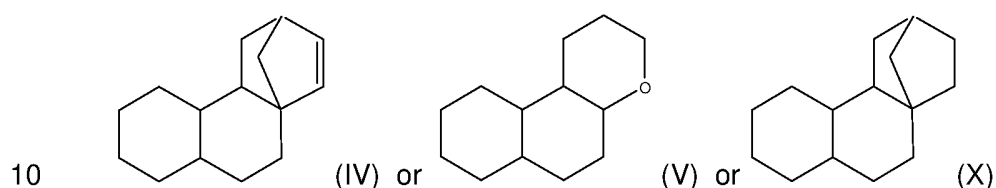
25 The sequence identity is preferably calculated as described herein below in the section "Sequence identity". A functional homologue of CfTPS4 is a polypeptide, which is also capable of catalysing at least one of reactions XXII, XXIII or XXIV described above.

TwTPS2

30

The invention involves use of a diTPS of class I. In one embodiment said diTPS of class I may be a TwTPS2 like diTPS.

In particular, said diTPS of class I may be a TwTPS2 like diTPS in embodiments of the invention, wherein the diterpene to be produced contains a tricyclic ring structure. For example said diTPS of class I may be a TwTPS2 like diTPS in embodiments of the invention, wherein the diterpene to be produced contains a core of any of the formulas IV, V or X:

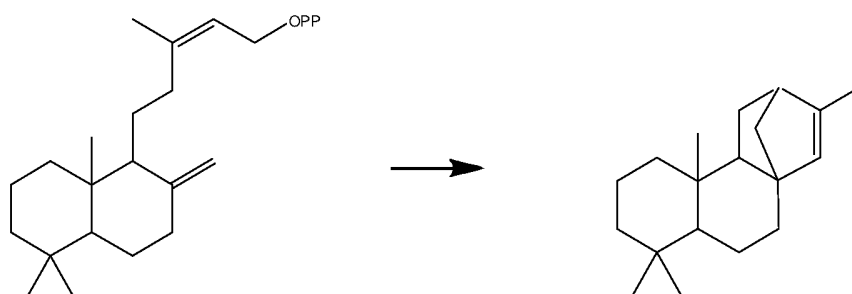


Dependent on the structure of the diterpene pyrophosphate intermediate then the diterpene containing a core of formula IV and V, may have different stereochemistry. In general the stereochemistry of the decalin core present in the diterpene pyrophosphate intermediate is maintained after the reaction catalysed by the TwTPS2 like diTPS.

The TwTPS2 like diTPS may be any enzyme capable of catalysing the reaction XXVI:

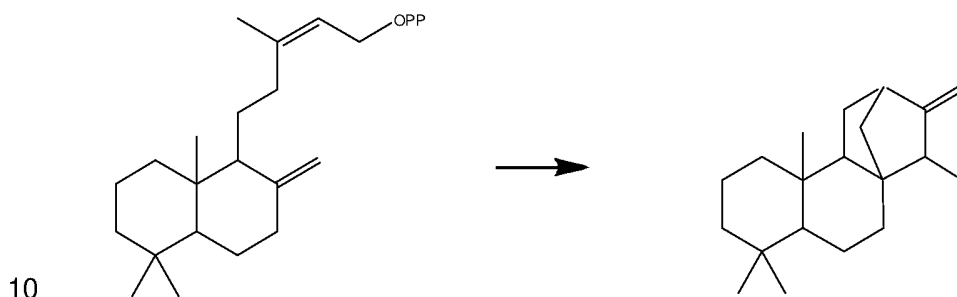
20 Diterpene pyrophosphate intermediate containing a decalin core structure \longrightarrow
Diterpene containing a core structure of formula IV or formula V or formula X

The TwTPS2 like diTPS may be any enzyme capable of catalysing conversion of a diterpene pyrophosphate intermediate to a diterpene containing a core of either formula IV or V. The TwTPS2 like diTPS may in particular be an enzyme capable of catalysing the reaction XIX:



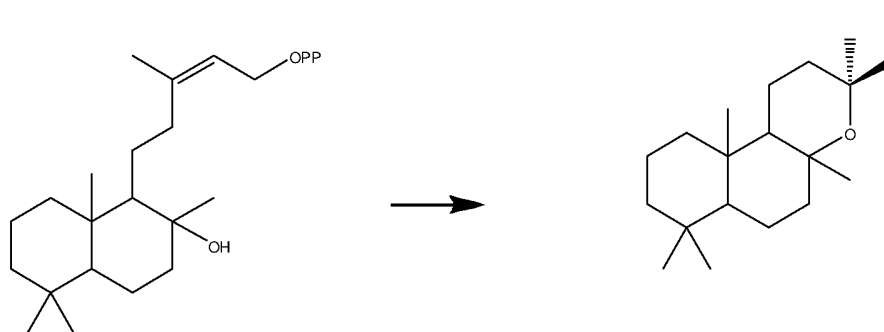
wherein OPP is diphosphate. During reaction XIX the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

The TwTPS2 like diTPS may in particular be an enzyme capable of catalysing the reaction XXVII:



wherein OPP is diphosphate. During reaction XIX the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

The TwTPS2 like diTPS may in particular be an enzyme capable of catalysing the reaction XX:



wherein OPP indicated diphosphate. During reaction XX the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

5

In one embodiment the TwTPS2 like diTPS may be a diterpene synthase from *Tripterygium Wilfordii*. In particular, the TwTPS2 like diTPS may be a TPS2 from *Tripterygium Wilfordii*. TPS2 from *Tripterygium Wilfordii* may also be referred to as TwTPS2. In particular, said TwTPS2 like diTPS may be a polypeptide of SEQ ID NO:14 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith.

15

The sequence identity is preferably calculated as described herein below in the section "Sequence identity". A functional homologue of TwTPS2 is a polypeptide, which is also capable of catalysing at least one of reactions, XIX, XX, XXVI or XXVII described above.

20

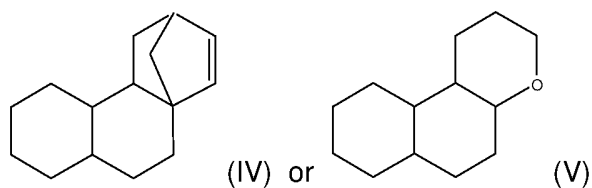
EpTPS1

The invention involves use of a diTPS of class I. In one embodiment said diTPS of class I may be an EpTPS1 like diTPS.

25

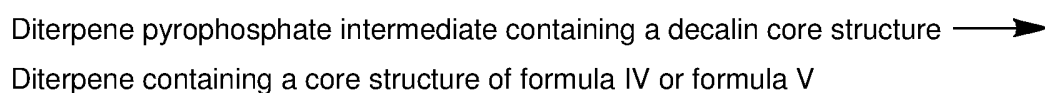
In particular, said diTPS of class I may be an EpTPS1 like diTPS in embodiments of the invention, wherein the diterpene to be produced contains a tricyclic ring structure. For example said diTPS of class I may be an EpTPS1 like diTPS in embodiments of the invention, wherein the diterpene to be produced contains a core of any of the formulas IV or V:

30

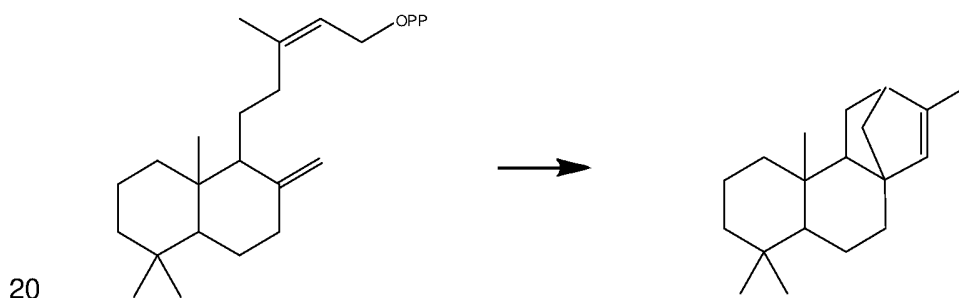


5 Dependent on the structure of the diterpene pyrophosphate intermediate then the diterpene containing a core of formula IV and V, may have different stereochemistry. In general the stereochemistry of the decalin core present in the diterpene pyrophosphate intermediate is maintained after the reaction catalysed by the EpTPS1 like diTPS.

10 The EpTPS1 like diTPS may be any enzyme capable of catalysing the reaction XVIII:



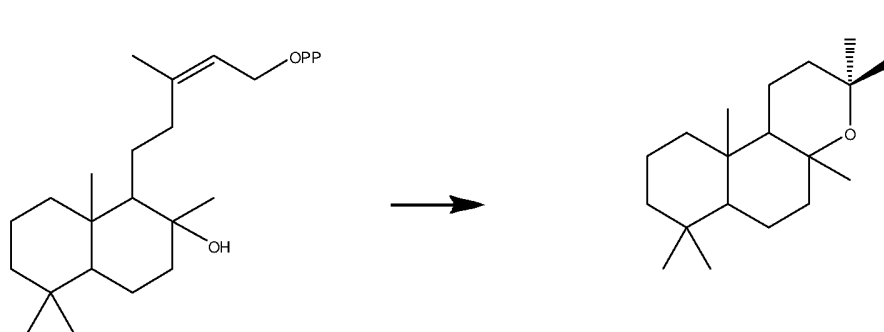
15 The EpTPS1 like diTPS may be any enzyme capable of catalysing conversion of a diterpene pyrophosphate intermediate to a diterpene containing a core of either formula IV or V. The EpTPS1 like diTPS may in particular be an enzyme capable of catalysing the reaction XIX:



wherein OPP is diphosphate. During reaction XIX the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

25

The EpTPS1 like diTPS may in particular be an enzyme capable of catalysing the reaction XX:



wherein OPP indicated diphosphate. During reaction XX the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

5

In one embodiment the EpTPS1 like diTPS may be a diterpene synthase from *Euphobia peplus*. In particular, the EpTPS1 like diTPS may be a TPS1 from *Euphobia peplus*. TPS1 from *Euphobia peplus* may also be referred to as EpTPS1. In particular, said EpTPS1 like diTPS may be a polypeptide of SEQ ID NO:15 or a functional
 10 homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith.

15

The sequence identity is preferably calculated as described herein below in the section "Sequence identity". A functional homologue of EpTPS1 is a polypeptide, which is also capable of catalysing at least one of reactions XVIII, XIX or XX described above.

20

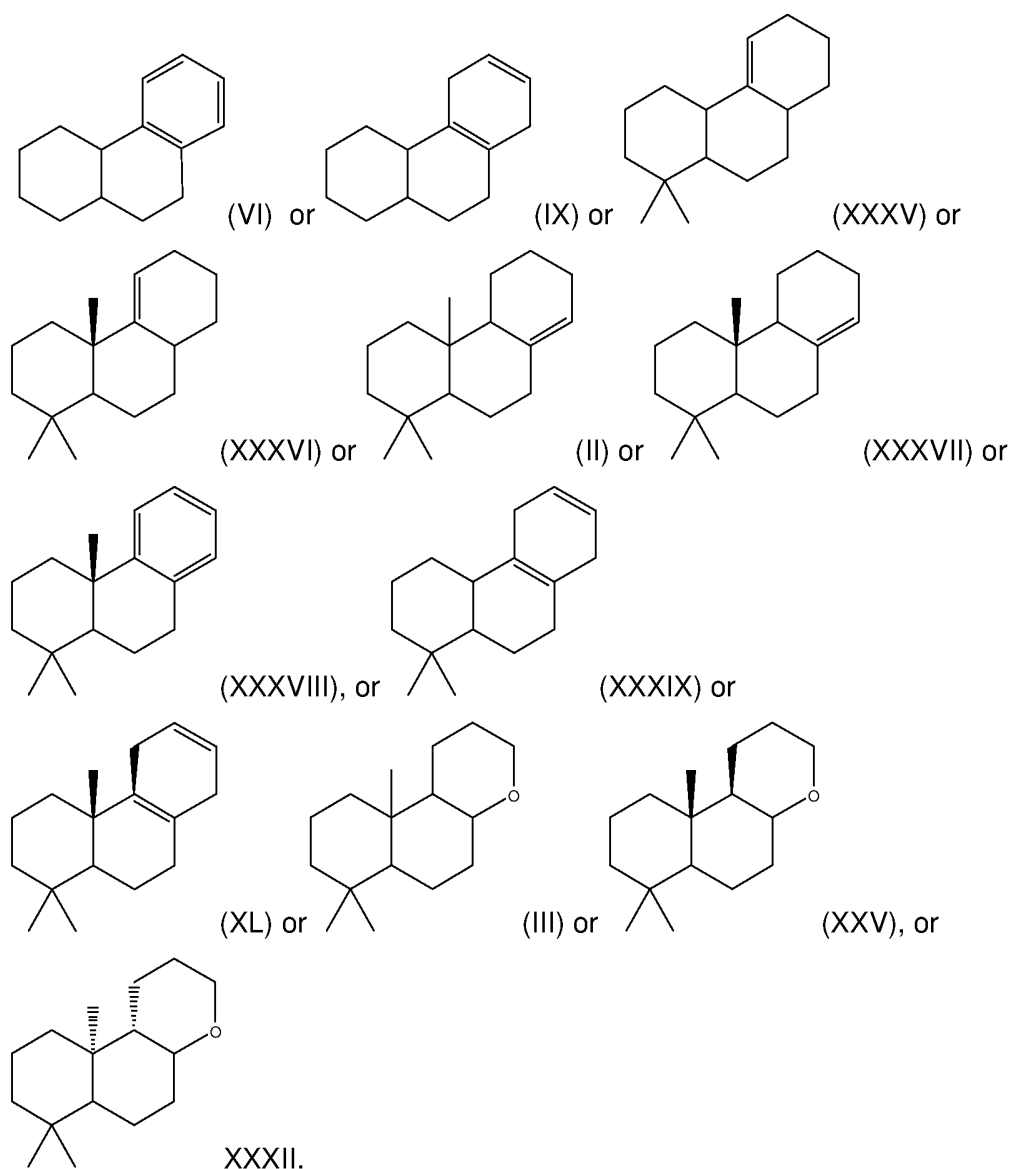
MvTPS5

The invention involves use of a diTPS of class I. In one embodiment said diTPS of class I may be a MvTPS5 like diTPS.

25

In particular, said diTPS of class I may be a MvTPS5 like diTPS in embodiments of the invention, wherein the diterpene to be produced contains a tricyclic ring structure. For example said diTPS of class I may be a MvTPS5 like diTPS in embodiments of the invention, wherein the diterpene to be produced contains a core of any of the formulas VI, IX, XXXV, XXXVI, II,XXXVII, XXXVIII, XXXIX, XL, III or XXXII:

30



5

Dependent on the structure of the diterpene pyrophosphate intermediate then the diterpene containing a core of formula VI, IX, XXXV, II, XXXIX or III, may have different stereochemistry. In general the stereochemistry of the decalin core present in the diterpene pyrophosphate intermediate is maintained after the reaction catalysed by the MvTPS5 like diTPS.

10

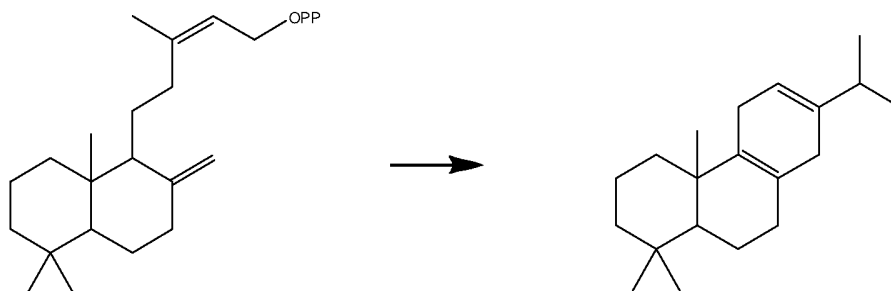
The MvTPS5 like diTPS may be any enzyme capable of catalysing the reaction XXIII:

15

Diterpene pyrophosphate intermediate containing a decalin core structure \longrightarrow

Diterpene containing a core structure of formula VI, IX, XXXV, XXXVI, II,XXXVII, XXXVIII, XXXIX, XL, III or XXXII.

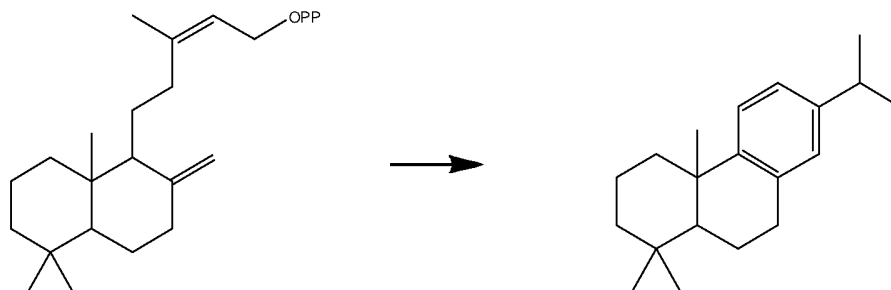
5 The MvTPS5 like diTPS may in particular be an enzyme capable of catalysing the reaction XXIV:



wherein OPP indicates diphosphate. During reaction XXIV the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

10

The MvTPS5 like diTPS may in particular be an enzyme capable of catalysing the reaction XXII:

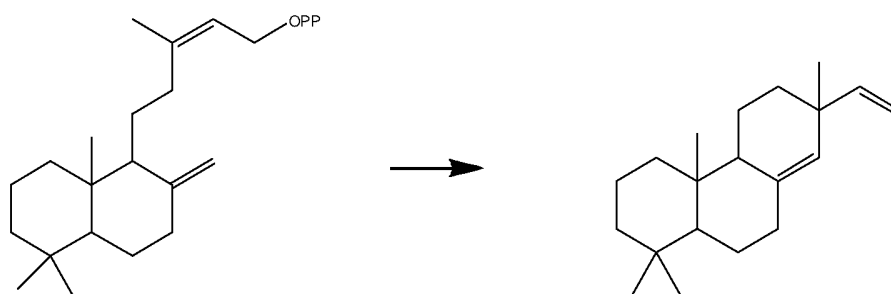


15

wherein OPP is diphosphate. During reaction XXII the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

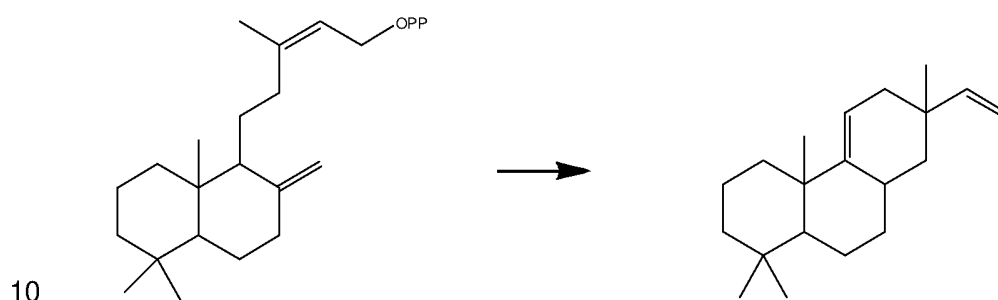
20

The MvTPS5 like diTPS may in particular be an enzyme capable of catalysing the reaction XXXI:



wherein OPP is diphosphate. During reaction XXXI the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting
5 diterpene pyrophosphate intermediate.

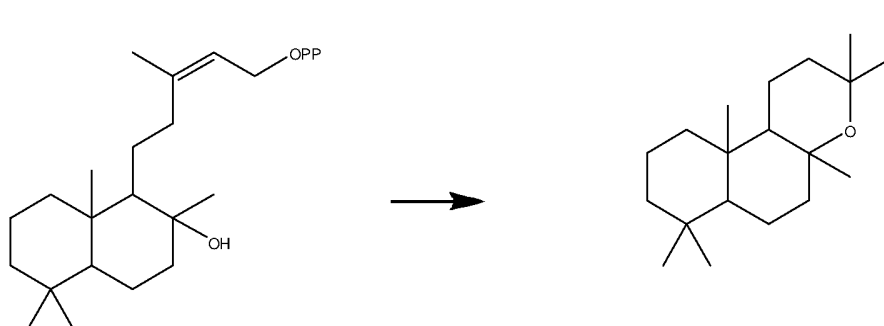
The MvTPS5 like diTPS may in particular be an enzyme capable of catalysing the reaction XXXII:



wherein OPP is diphosphate. During reaction XXXII the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting
10 diterpene pyrophosphate intermediate.

15

The MvTPS5 like diTPS may also be an enzyme catalysing the reaction X:



wherein OPP indicates diphosphate. During reaction X the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

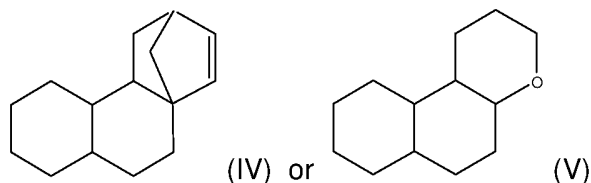
- 5 In one embodiment the MvTPS5 like diTPS may be a diterpene synthase from *Marrubium vulgare*. In particular, the MvTPS5 like diTPS may be a TPS5 from *Marrubium vulgare*. TPS5 from *Marrubium vulgare* may also be referred to as MvTPS5. In particular, said MvTPS5 like diTPS may be a polypeptide of SEQ ID NO:18 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example
- 10 at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith.
- 15 The sequence identity is preferably calculated as described herein below in the section "Sequence identity". A functional homologue of MvTPS5 is a polypeptide, which is also capable of catalysing at least one of reactions XXII, XXIII or XXIV described above.

20 **CfTPS14**

The invention involves use of a diTPS of class I. In one embodiment said diTPS of class I may be an CfTPS14 like diTPS.

- 25 In particular, said diTPS of class I may be an CfTPS14 like diTPS in embodiments of the invention, wherein the diterpene to be produced contains a tricyclic ring structure. For example said diTPS of class I may be an CfTPS14 like diTPS in embodiments of the invention, wherein the diterpene to be produced contains a core of any of the formulas IV or V:

30



Dependent on the structure of the diterpene pyrophosphate intermediate then the diterpene containing a core of formula IV and V, may have different stereochemistry. In general the stereochemistry of the decalin core present in the diterpene pyrophosphate intermediate is maintained after the reaction catalysed by the CfTPS14 like diTPS.

5

The CfTPS14 like diTPS may be any enzyme capable of catalysing the reaction XVIII:

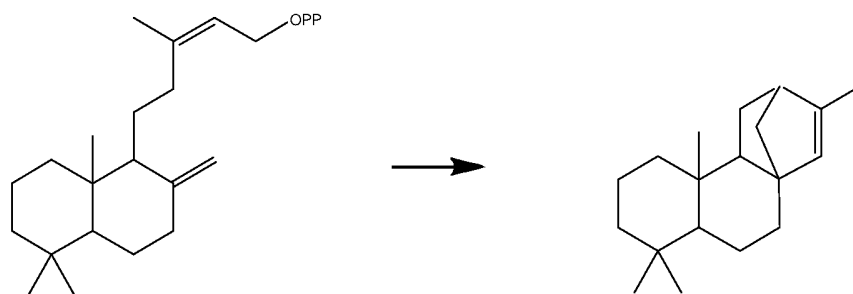
Diterpene pyrophosphate intermediate containing a decalin core structure \longrightarrow

Diterpene containing a core structure of formula IV or formula V

10

The CfTPS14 like diTPS may be any enzyme capable of catalysing conversion of a diterpene pyrophosphate intermediate to a diterpene containing a core of either formula IV or V. The CfTPS14 like diTPS may in particular be an enzyme capable of catalysing the reaction XIX:

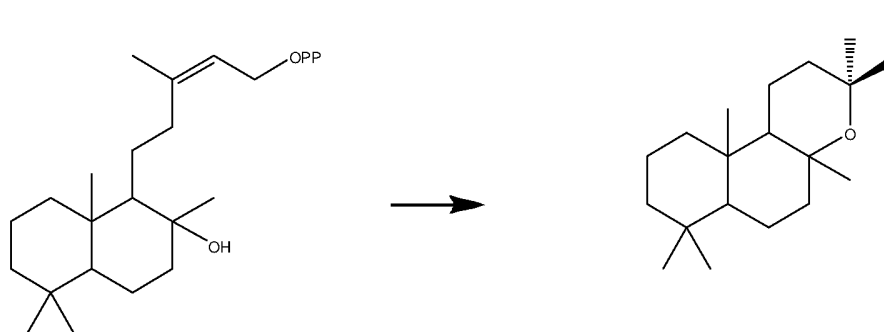
15



wherein OPP is diphosphate. During reaction XIX the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

20

The CfTPS14 like diTPS may in particular be an enzyme capable of catalysing the reaction XX:



wherein OPP indicated diphosphate. During reaction XX the produced diterpene will in general maintain the stereochemistry around the decalin core found in the starting diterpene pyrophosphate intermediate.

5

In one embodiment the CfTPS14 like diTPS may be a diterpene synthase from *Coleus forskohlii*. In particular, the CfTPS14 like diTPS may be a TPS14 from *Coleus forskohlii*. TPS14 from *Coleus forskohlii* may also be referred to as CfTPS14. In particular, said CfTPS14 like diTPS may be a polypeptide of SEQ ID NO:16 or a functional homologue thereof sharing at least 70%, such as at least 80%, for example at least 75%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99%, such as 100% sequence identity therewith.

15

The sequence identity is preferably calculated as described herein below in the section "Sequence identity". A functional homologue of CfTPS14 is a polypeptide, which is also capable of catalysing at least one of reactions XVIII, XIX or XX described above.

20 Additional recombinant modifications

The host organisms according to the present invention may also be recombinantly modified in addition to comprising the heterologous nucleic acids encoding a diTPS of class I and a diTPS of class II as described herein.

25

For example the host organism may be modified to increase the pool of GGPP. As described herein elsewhere, GGPP is the starting compound for production of diterpenes. Thus, if the host organism is modified to increase the pool of GGPP, then

frequently, the host organism will be capable of producing increased amounts of diterpene.

5 Various methods for increasing the pool of GGPP are well known in the art. These includes methods of reducing the activity of enzymes reducing the level of GGPP.

In one embodiment the pool of GGPP is increased by expression of one or more enzymes involved in synthesis of GGPP.

10 Thus, it may be preferred that the host organism comprises a heterologous nucleic acid encoding GGPP synthase (GGPPS). Said GGPPS may be any GGPPS, e.g. BTS1 of *S. cerevisiae*.

15 In particular, the GGPPS may be the GGPPS described by Zhou, Y. J., W. Gao, Q. Rong, G. Jin, H. Chu, W. Liu, W. Yang, Z. Zhu, G. Li, G. Zhu, L. Huang and Z. K. Zhao (2012). "Modular Pathway Engineering of Diterpenoid Synthases and the Mevalonic Acid Pathway for Miltiradiene Production." Journal of the American Chemical Society **134**(6): 3234-3241.

20 Accordingly, the host organism may express a fusion of SmCPS and SmKSL, and/or a fusion of BTS1 (GGPP synthase) and ERG20 (farnesyl diphosphate synthase) as described in Zhou et al., 2012.

25 The host organism may also comprise a heterologous nucleic acid encoding a GGPPS from a plant, e.g. from *Coleus forskohlii*. Thus, in one embodiment the host organism comprises:

- a) a heterologous nucleic acid encoding *Coleus forskohlii* deoxyxylulose 5-phosphate synthase (CfDXS) of SEQ ID NO:26 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%,
30 such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith and/or
- b) a heterologous nucleic acid encoding *Coleus forskohlii* geranylgeranylpyrophosphate synthase (CfGGPPs) of SEQ ID NO:27 or a functional homologue of any of the aforementioned sharing at least 70%,
35 such as at least 80%, such as at least 85%, such as at least 90%, such as

at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

5 **Production of kolavelool**

It is one aspect of the invention to provide methods for producing kolavelool. In particular, the invention provides methods for producing kolavelool, said methods comprising the steps of:

10

a) providing a host organism comprising

- I. a heterologous nucleic acid encoding a diTPS of class II, which is an CLPP like type diTPS; and
- II. A heterologous nucleic acid encoding diTPS of class I,

15

- b) Incubating said host organism in the presence of geranylgeranyl pyrophosphate (GGPP) under conditions allowing growth of said host organism;
- c) Optionally isolating kolavelool from the host organism.

20

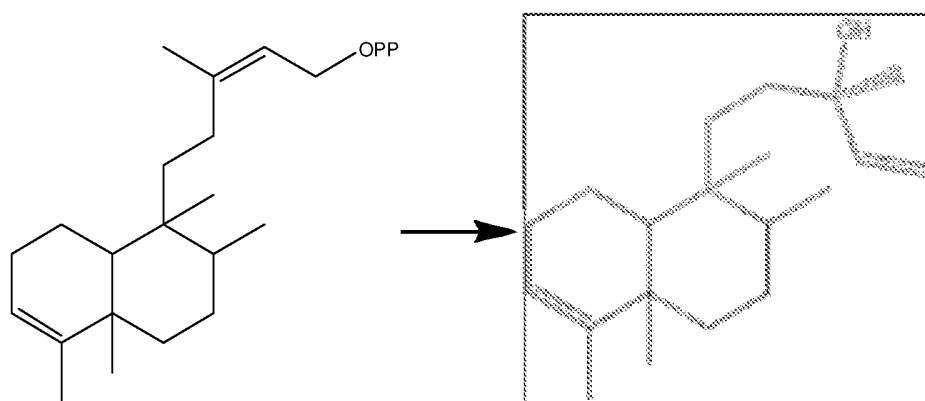
Said host organism may for example be any of the host organisms described herein in the section "Host organism".

25

Said CLPP type diTPS may be any of the CLPP type diTPS described herein in the section "LPP type diTPS". In particular the LPP type diTPS may be TwTPS14/28 of SEQ ID NO:8 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith. Said functional homologue is preferably an enzyme capable of catalysing reaction XXXV.

30

The diTPS of class I may be any diTPS of class I, such as any of the diTPS of class I described herein. In particular, said diTPS of class I may be a diTPS of class I capable of catalysing the reaction XXXVII:



In one preferred embodiment of the invention, the diTPS of class I may in embodiment be a SsSCS like diTPS, for example any of the SsSCS like diTPS described herein in the section "ScSCS". In particular the SsSCS like diTPS may be SsSCS of SEQ ID NO:11 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Sequence identity

A high level of sequence identity indicates likelihood that the first sequence is derived from the second sequence. Amino acid sequence identity requires identical amino acid sequences between two aligned sequences. Thus, a candidate sequence sharing 80% amino acid identity with a reference sequence, requires that, following alignment, 80% of the amino acids in the candidate sequence are identical to the corresponding amino acids in the reference sequence. Identity according to the present invention is determined by aid of computer analysis, such as, without limitations, the ClustalW computer alignment program (Higgins D., Thompson J., Gibson T., Thompson J.D., Higgins D.G., Gibson T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673-4680), and the default parameters suggested therein. The ClustalW software is available from as a ClustalW WWW Service at the European Bioinformatics Institute <http://www.ebi.ac.uk/clustalw> or via the software BioEdit. Using this program with its default settings, the mature (bioactive) part of a query and a reference polypeptide are aligned. The number of fully conserved residues are counted and divided by the length of the reference polypeptide.

Thus, sequence identity is calculated over the entire length of the reference polypeptide.

The ClustalW algorithm may similarly be used to align nucleotide sequences.

5 Sequence identities may be calculated in a similar way as indicated for amino acid sequences.

In one important embodiment, the cell of the present invention comprises a nucleic acid sequence coding, as define herein.

10

Heterologous nucleic acid

The term "heterologous nucleic acid" as used herein refers to a nucleic acid sequence, which has been introduced into the host organism, wherein said host does not
15 endogenously comprise said nucleic acid. For example, said heterologous nucleic acid may be introduced into the host organism by recombinant methods. Thus, the genome of the host organism has been augmented by at least one incorporated heterologous nucleic acid sequence. It will be appreciated that typically the genome of a recombinant host described herein is augmented through the stable introduction of one or more
20 heterologous nucleic acids encoding one or more diTPS's.

Suitable host organisms include microorganisms, plant cells, and plants, and may for example be any of the host organisms described herein below in the section "Host organism".

25

In general the heterologous nucleic acid encoding a polypeptide (also referred to as "coding sequence" in the following) is operably linked in sense orientation to one or more regulatory regions suitable for expressing the polypeptide. Because many microorganisms are capable of expressing multiple gene products from a polycistronic
30 mRNA, multiple polypeptides can be expressed under the control of a single regulatory region for those microorganisms, if desired. A coding sequence and a regulatory region are considered to be operably linked when the regulatory region and coding sequence are positioned so that the regulatory region is effective for regulating transcription or translation of the sequence. Typically, the translation initiation site of the translational

reading frame of the coding sequence is positioned between one and about fifty nucleotides downstream of the regulatory region for a monocistronic gene.

5 “Regulatory region” refers to a nucleic acid having nucleotide sequences that influence transcription or translation initiation and rate, and stability and/or mobility of a transcription or translation product. Regulatory regions include, without limitation, promoter sequences, enhancer sequences, response elements, protein recognition sites, inducible elements, protein binding sequences, 5′ and 3′ untranslated regions (UTRs), transcriptional start sites, termination sequences, polyadenylation sequences, 10 introns, and combinations thereof. A regulatory region typically comprises at least a core (basal) promoter. A regulatory region also may include at least one control element, such as an enhancer sequence, an upstream element or an upstream activation region (UAR). A regulatory region is operably linked to a coding sequence by positioning the regulatory region and the coding sequence so that the regulatory region 15 is effective for regulating transcription or translation of the sequence. For example, to operably link a coding sequence and a promoter sequence, the translation initiation site of the translational reading frame of the coding sequence is typically positioned between one and about fifty nucleotides downstream of the promoter. A regulatory region can, however, be positioned at further distance, for example as much as about 20 5,000 nucleotides upstream of the translation initiation site, or about 2,000 nucleotides upstream of the transcription start site.

The choice of regulatory regions to be included depends upon several factors, including the type of host organism. It is a routine matter for one of skill in the art to 25 modulate the expression of a coding sequence by appropriately selecting and positioning regulatory regions relative to the coding sequence. It will be understood that more than one regulatory region may be present, *e.g.*, introns, enhancers, upstream activation regions, transcription terminators, and inducible elements.

30 It will be appreciated that because of the degeneracy of the genetic code, a number of nucleic acids can encode a particular polypeptide; *i.e.*, for many amino acids, there is more than one nucleotide triplet that serves as the codon for the amino acid. Thus, codons in the coding sequence for a given polypeptide can be modified such that optimal expression in a particular host organisms obtained, using appropriate codon 35 bias tables for that host (*e.g.*, microorganism). Nucleic acids may also be optimized to

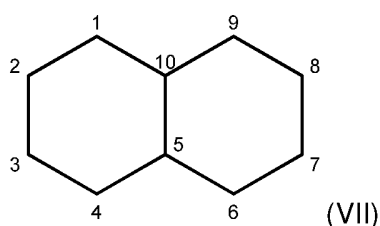
a GC-content preferable to a particular host, and/or to reduce the number of repeat sequences. As isolated nucleic acids, these modified sequences can exist as purified molecules and can be incorporated into a vector or a virus for use in constructing modules for recombinant nucleic acid constructs.

5

Diterpene pyrophosphate intermediate

The term "decalin" as used herein refers to a compound of the formula VII:

10



The numbering of carbon atoms provided in formula VII is adhered to throughout this description.

15

A compound containing or comprising a "decalin core" as used herein refers to a compound comprising above mentioned structure of formula VII, wherein each of the carbon atoms numbered 1 to 10 may be substituted with one or two substituents. It is possible that two of said substituents are fused to form a ring, and thus compound containing or comprising decalin may contain 3 or more rings.

20

The term "diterpene pyrophosphate intermediate" as used herein refers to a compound, which is the product of bicyclisation of GGPP in a reaction catalysed by a diTPS class II enzyme. The diterpene pyrophosphate intermediate according to the invention contains a decalin core, and comprises a pyrophosphate group.

25

It is preferred that the diterpene pyrophosphate intermediate of the invention is a compound containing a decalin core, which is substituted at one of more positions with substituents selected from the group consisting of alkyl, alkenyl and hydroxyl, wherein one of said alkyl or alkenyl is substituted with O-pyrophosphate.

30

The terms "diphosphate" and "pyrophosphate" are used interchangeably herein. The abbreviation "OPP", "-OPP" or "PPO-" as used herein refers to diphosphate.

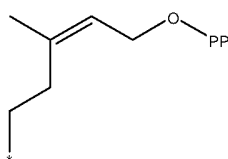
5 The term "alkyl" as used herein refers to a saturated, straight or branched hydrocarbon chain. The hydrocarbon chain preferably contains of from one to eighteen carbon atoms (C_{1-18} -alkyl), more preferred of from one to six carbon atoms (C_{1-6} -alkyl), including methyl, ethyl, propyl, isopropyl, butyl, isobutyl, secondary butyl, tertiary butyl, pentyl, isopentyl, neopentyl, tertiary pentyl, hexyl and isohexyl.

10 The term "alkenyl" as used herein refers to a saturated, straight or branched hydrocarbon chain containing at least one double bond. Alkenyl may preferably be any of the alkyls described above containing one or more double bonds.

15 In particular, the diterpene pyrophosphate intermediate of the invention is a compound containing a decalin core, wherein said decalin is

- i. substituted at the 4 position with one or two alkyl, such as with two alkyl, wherein said alkyl for example may be C_{1-3} , alkyl, for example said alkyl may be methyl;
- ii. substituted at the 8 position with one or two substituents individually
20 selected from the group consisting of alkyl, hydroxyl and alkenyl, wherein said alkyl for example may be C_{1-3} alkyl, for example said alkyl may be methyl, and said alkenyl may be C_{1-3} alkenyl, for example said alkenyl may be =C;
- iii. substituted at the 9 position with alkenyl-O-PP, wherein said alkenyl for
25 example may be branched C4-8-alkenyl, such as branched C5-7-alkenyl, for example branched C6-alkenyl; and
- iv. substituted at the 10 position with alkyl, wherein said alkyl for example may be C_{1-3} , alkyl, for example said alkyl may be methyl.

30 In particular, the substituent at the 9 position may be alkenyl of formula VIII:



wherein the asterisk indicates the point of attachment to the decalin core.

It is also preferred that the stereochemistry around substituents 9 and 10 is predetermined. Thus, said diterpene pyrophosphate intermediate may contain a decalin core substituted as indicated above, wherein the substitutions at the 9 and 10 positions are (9R, 10R), (9S,10S), (9S, 10R) or (9R, 10S), for example the substitutions
5 at the 9 and 10 positions are (9R, 10R), (9S,10S) or (9S, 10R).

In preferred embodiments, the diterpene pyrophosphate intermediate may be any of the diterpene pyrophosphate intermediates shown in figure 3, i.e. the diterpene pyrophosphate intermediate may be selected from the group consisting of (9R,10R)-
10 copalyl diphosphate, (9S,10S)-copalyl diphosphate, labda-13-en-8-ol diphosphate and (9S, 10R)-copalyl diphosphate.

Diterpenes

15 The term "diterpene" as used herein refers to a compound derived or prepared from four isoprene units. A diterpene according to the invention is a C₂₀- molecule consisting of 20 carbon atoms, up to three oxygen atoms and hydrogen atoms.

The diterpene typically contains one or more ring structures, such as one or more
20 monocyclic, bicyclic, tricyclic or tetracyclic ring structure(s). The diterpene may contain one or more double bonds. Frequently, a diterpene according to the invention contains at least one double bond and often they contain in the range of 1 to 3 double bonds.

The diterpene may comprise up to three oxygen atom, although it is also possible that
25 the diterpene contains no oxygen and consists solely of carbon and hydrogen atoms. The oxygen atom are generally present in the form of hydroxyl groups, or part of a ring structure.

The term "diterpenoid" refers to a diterpene, which has been functionalised by addition
30 of one or more functional groups.

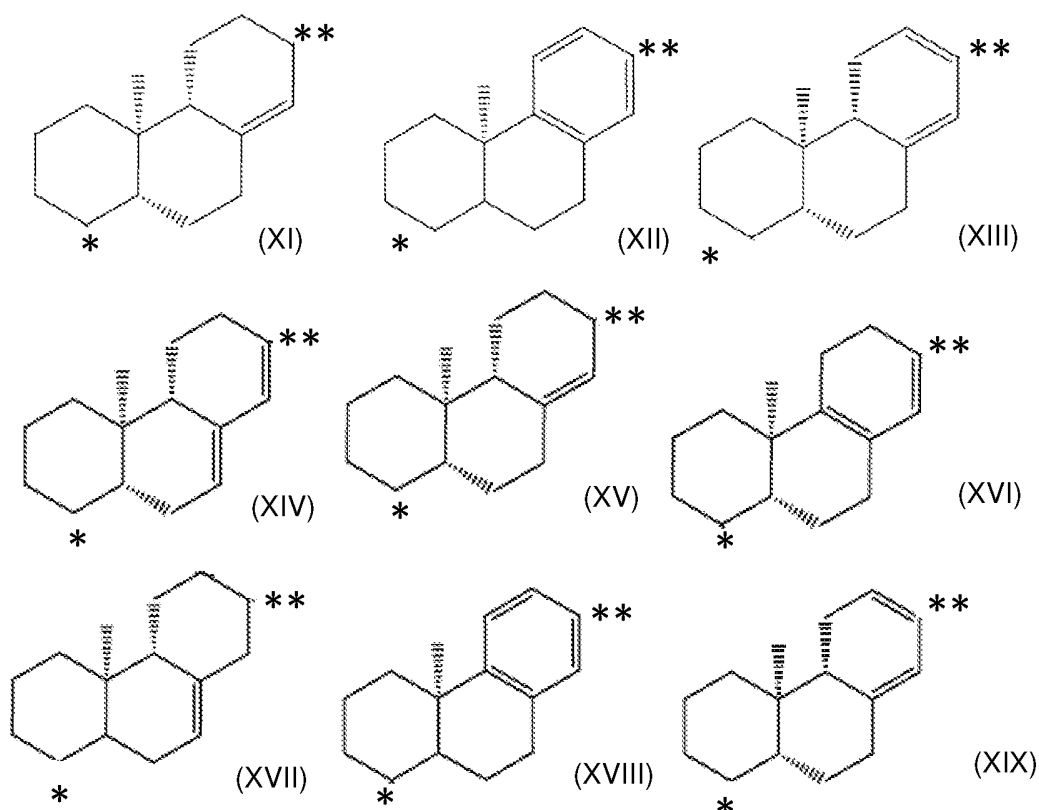
In principle, the methods of the invention can be used to produce any diterpene by selecting an appropriate combination of diTPS of class II and diTPS of class I.

In one preferred embodiment the diterpene to be produce is a C₂₀-molecule containing a decalin core structure.

5 As used herein the term “containing a core structure of formula ” or the term “containing a core of formula” refers to a molecule containing a structure of the indicated formula, wherein said structure may be substituted at one or more positions. The term “substituted” as used herein in relation to organic compounds refer to one hydrogen being substituted with another group or atom.

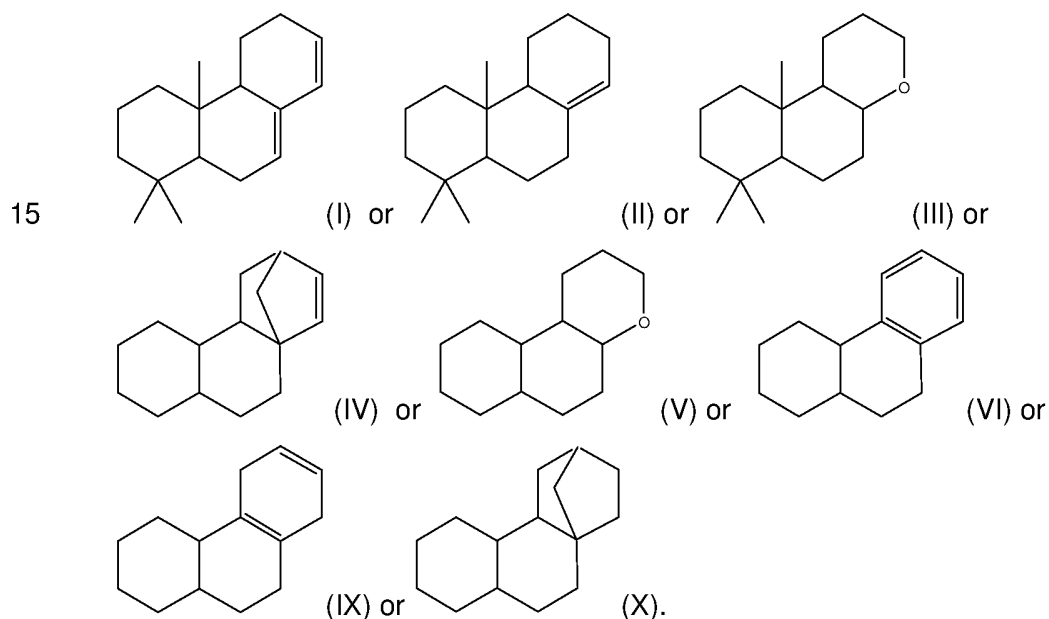
10 Said decalin may be substituted at one or more positions, and it is also contained within the invention that two substituents are fused, thus leading to a tricyclic or higher cyclic structure.

15 In particular, the diterpene to be produced by the methods of the present invention may be a C₂₀-molecule containing a core structure of one of following formulas XI, XII, XIII, XIV, XV, XVI, XVII, XVIII or XIX:



The diterpene containing a core structure of any of formulas XI, XII, XIII, XIV, XV, XVI, XVII, XVIII or XIX, may be a C_{20} -molecule consisting of the formulas XI, XII, XIII, XIV, XV, XVI, XVII, XVIII or XIX substituted at one or more positions. In particular, said
 5 diterpene may be a C_{20} -molecule substituted at the position marked by * with one or two alkyl, such as one or two C_{1-3} -alkyl, such as with one or two methyl groups. In addition said diterpene may be substituted at the position marked by ** with one or two groups individually selected from alkyl and alkenyl. Said alkyl may for example be C_{1-6} -alkyl, such as C_{1-3} -alkyl, for example isopropyl or methyl. Said alkenyl may be C_{1-6}
 10 alkenyl, such as C_{2-4} -alkenyl, such as C_{2-3} -alkenyl.

In preferred embodiments of the invention the diterpene to be produced may be a C_{20} -molecule containing a core structure of one of following formulas I, II, III, IV, V, VI, IX or X:



The diterpene containing a core structure of any of formulas I, II, III, IV, V, VI, IX or X,
 20 may be a C_{20} -molecule consisting of the formulas I, II, III, IV, V, VI, IX or X substituted at one or more positions, for example by one or more groups selected from the group consisting of:

- c) alkyl, such as C_{1-6} -alkyl, for example C_{1-3} , wherein said alkyl may be linear or branched, for example alkyl may be isopropyl or methyl
- 25 d) alkenyl, such as C_{1-6} alkenyl, such as C_{2-4} -alkenyl, such as C_{2-3} -alkenyl
- e) hydroxyl

In particular said diterpene containing a core structure of any of formulas I, II, III, IV, V, VI, IX or X, may be a C₂₀-molecule substituted

- 5 a) at the position corresponding to the 4 position of decalin with one or two alkyl, such as one or two C₁₋₃-alkyl, such as with one or two methyl groups, for example with two methyl; and/or
- b) at the position corresponding to the 10 position of decalin with alkyl, such as with C₁₋₃-alkyl, such as with methyl; and/or
- 10 c) at the position corresponding to the position marked by ** in relations to formulas XI-XIX, with one or two groups individually selected from alkyl and alkenyl. Said alkyl may for example be C₁₋₆-alkyl, such as C₁₋₃-alkyl, for example isopropyl or methyl. Said alkenyl may be C₁₋₆ alkenyl, such as C₂₋₄-alkenyl, such as C₂₋₃-alkenyl; and/or
- d) hydroxyl.

15

The diterpene to be produced may also be a C₂₀-molecule consisting of 20 carbon atoms, up to three oxygen atoms and hydrogen atoms, and which contains a core structure of any of formulas I, II, III, IV, VI, X, XXII, XXIII, XXIV, XXV, XXVI, XXVII, XXVIII, XXIX, XXX, XXXI, XXXII, XXXIII, XXXIV, XXXV, XXXVI, XXXVII, XXXVIII, XXXIX, XL and/or XLI.

20

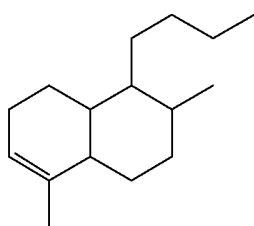
The diterpene to be produced may also be a C₂₀-molecule consisting of 20 carbon atoms, up to three oxygen atoms and hydrogen atoms, and which contains a core structure of any of formulas I, II, IV, VI, X, XXII, XXIII, XXIV, XXVI, XXVII, XXVIII, XXIX, XXX, XXXI, XXXIII, XXXIV, XXXV, XXXVI, XXXVII, XXXVIII, XXXIX, XL and/or XLI.

25

The structure of the formulas I, II, III, IV, VI, X, XXII, XXIII, XXIV, XXV, XXVI, XXVII, XXVIII, XXIX, XXX, XXXI, XXXII, XXXIII, XXXIV, XXXV, XXXVI, XXXVII, XXXVIII, XXXIX, XL and XLI are as indicated herein above.

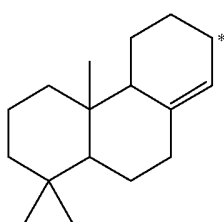
30

In one embodiment the diterpene is a C₂₀-molecule containing a core of formula XXXIII:

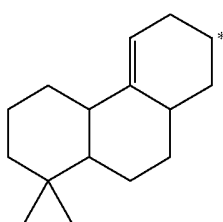


(XXXIII). Said diterpene may in particular contain a core of formula XXXIII substituted with alkyl, alkenyl and/or hydroxyl, preferably substituted with methyl, =CH₂ and hydroxyl.

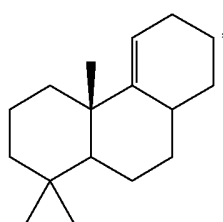
In another embodiment the diterpene is a C₂₀-molecule containing a core of any of
 5 formulas II, XXXV, XXXVI and/or XXXVII:



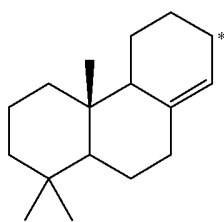
(II), or



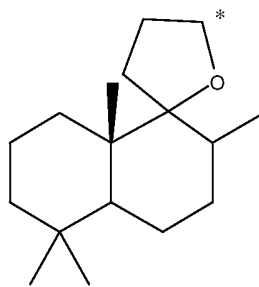
(XXXV) or



(XXXVI) or



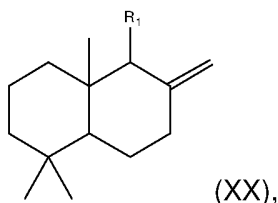
(XXXVII) or or



(XLI), wherein said core may be

substituted with one or more alkyl or alkenyl. In particular, the position marked by
 10 asterisk may be substituted with one or two substituents selected from the group consisting of C₁₋₂-alkyl and C₁₋₂-alkenyl, preferably the position marked by asterisk may be substituted with one methyl group and ethenyl group.

In one embodiment, said diterpene to be produced is a C₂₀-molecule containing a
 15 decalin substituted at the 10 position with C₅-alkenyl chain, which optionally may be substituted with a hydroxyl and/or a methyl group and/or =C. For example, said diterpene may be a C₂₀-molecule of the formula XX:

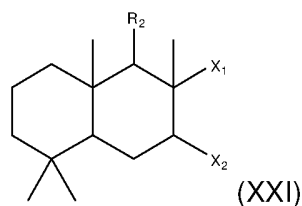


(XX),

wherein R_1 is a C_5 -alkenyl substituted with methyl and/or hydroxyl. Preferably, R_1 is C_5 -alkenyl containing one or two double bonds. When R_1 is alkenyl containing one double bond, said alkenyl is preferably substituted with hydroxyl and methyl. When R_1 is alkenyl containing two double bonds, said alkenyl is preferably substituted with methyl.

5

For example, said diterpene may be a C_{20} -molecule of the formula XXI:



wherein R_2 is a C_5 -alkenyl substituted with methyl and/or hydroxyl or with $=C$, and X_1 is either $-OH$ or methyl, and X_2 is either $-H$ or $-OH$, wherein one and only one of X_1 and X_2 is $-OH$. Preferably, R_2 is C_5 -alkenyl containing one or two double bonds. When R_2 is alkenyl containing one double bond, said alkenyl is preferably substituted with hydroxyl and methyl or with $=C$. When R_2 is alkenyl containing two double bonds, said alkenyl is preferably substituted with methyl.

15

It is also comprised within the invention that the diterpene is the product of any of the reactions VII to XIX described herein above.

In particular, the diterpene may be any of the compounds 1 to 47 shown in figure 2 and/or Table 1.

20

It is preferred that the diterpene to be produced is not 13R-manoyl oxide.

Host organism

25

The host organism to be used with the methods of the invention, may be any suitable host organism containing

a heterologous nucleic acid encoding a diTPS of class II, which may be any of diTPS of class II described herein in any of the sections "diTPS of class II", "syn-CPP type diTPS", "ent-CPP type diTPS", "(+)-CPP type diTPS", "LPP type diTPS", and "LPP like type diTPS"; and

30

a heterologous nucleic acid encoding a diTPS of class I, which may be any of diTPS of class I described herein in any of the sections "diTPS of class I", "EpTPS8", "EpTPS23", "SsSCS", "CfTPS3", "CfTPS4", "MvTPS5", "TwTPS2", "EpTPS1", and "CfTPS14".

5

Suitable host organisms include microorganisms, plant cells, and plants.

The microorganism can be any microorganism suitable for expression of heterologous nucleic acids. In one embodiment the host organism of the invention is a eukaryotic cell. In another embodiment the host organism is a prokaryotic cell. In a preferred embodiment, the host organism is a fungal cell such as a yeast or filamentous fungus. In particular the host organism may be a yeast cell.

10

In a further embodiment the yeast cell is selected from the group consisting of *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Yarrowia lipolytica*, *Candida glabrata*, *Ashbya gossypii*, *Cyberlindnera jadinii*, and *Candida albicans*.

15

In general, yeasts and fungi are excellent microorganism to be used with the present invention. They offer a desired ease of genetic manipulation and rapid growth to high cell densities on inexpensive media. For instance yeasts grow on a wide range of carbon sources and are not restricted to glucose. Thus, the microorganism to be used with the present invention may be selected from the group of yeasts described below:

20

Arxula adeninivorans (*Blastobotrys adeninivorans*) is a dimorphic yeast (it grows as a budding yeast like the baker's yeast up to a temperature of 42 °C, above this threshold it grows in a filamentous form) with unusual biochemical characteristics. It can grow on a wide range of substrates and can assimilate nitrate. It has successfully been applied to the generation of strains that can produce natural plastics or the development of a biosensor for estrogens in environmental samples.

25

Candida boidinii is a methylotrophic yeast (it can grow on methanol). Like other methylotrophic species such as *Hansenula polymorpha* and *Pichia pastoris*, it provides an excellent platform for the production of heterologous proteins. Yields in a multigram range of a secreted foreign protein have been reported. A computational method, IPRO, recently predicted mutations that experimentally switched the cofactor specificity of *Candida boidinii* xylose reductase from NADPH to NADH. Details on how to

30

35

download the software implemented in Python and experimental testing of predictions are outlined in the following paper.

5 *Hansenula polymorpha* (*Pichia angusta*) is another methylotrophic yeast (see *Candida boidinii*). It can furthermore grow on a wide range of other substrates; it is thermo-tolerant and can assimilate nitrate (see also *Kluyveromyces lactis*). It has been applied to the production of hepatitis B vaccines, insulin and interferon alpha-2a for the treatment of hepatitis C, furthermore to a range of technical enzymes.

10 *Kluyveromyces lactis* is a yeast regularly applied to the production of kefir. It can grow on several sugars, most importantly on lactose which is present in milk and whey. It has successfully been applied among others to the production of chymosin (an enzyme that is usually present in the stomach of calves) for the production of cheese. Production takes place in fermenters on a 40,000 L scale.

15 *Pichia pastoris* is a methylotrophic yeast (see *Candida boidinii* and *Hansenula polymorpha*). It provides an efficient platform for the production of foreign proteins. Platform elements are available as a kit and it is worldwide used in academia for the production of proteins. Strains have been engineered that can produce complex human
20 N-glycan (yeast glycans are similar but not identical to those found in humans).

Saccharomyces cerevisiae is the traditional baker's yeast known for its use in brewing and baking and for the production of alcohol. As protein factory it has successfully been applied to the production of technical enzymes and of pharmaceuticals like insulin and
25 hepatitis B vaccines. Also it has been useful for production of terpenoids.

Yarrowia lipolytica is a dimorphic yeast (see *Arxula adeninivorans*) that can grow on a wide range of substrates. It has a high potential for industrial applications.

30 In another embodiment the host organism is a microalgae such as *Chlorella* and *Prototheca*.

 In another embodiment of the invention the host organism is a filamentous fungus, for example *Aspergillus*.

35

In further yet another embodiment the host organism is a plant cell. The host organism may be a cell of a higher plant, but the host organism may also be cells from organisms not belonging to higher plants for example cells from the moss *Physcomitrella patens*.

- 5 In another embodiment the host organism is a mammalian cell, such as a human, feline, porcine, simian, canine, murine, rat, mouse or rabbit cell.

- As mentioned, the host organism can also be a prokaryotic cell such as a bacterial cell. If the host organism is a prokaryotic cell the cell may be selected from, but not limited to *E. coli*, *Corynebacterium*, *Bacillus*, *Pseudomonas* and *Streptomyces* cells.
- 10

The host organism may also be a plant.

- A plant or plant cell can be transformed by having a heterologous nucleic acid integrated into its genome, *i.e.*, it can be stably transformed. Stably transformed cells typically retain the introduced nucleic acid with each cell division. A plant or plant cell can also be transiently transformed such that the recombinant gene is not integrated into its genome. Transiently transformed cells typically lose all or some portion of the introduced nucleic acid with each cell division such that the introduced nucleic acid cannot be detected in daughter cells after a certain number of cell divisions. Both transiently transformed and stably transformed transgenic plants and plant cells can be useful in the methods described herein.
- 15
- 20

- Plant cells comprising a heterologous nucleic acid used in methods described herein can constitute part or all of a whole plant. Such plants can be grown in a manner suitable for the species under consideration, either in a growth chamber, a greenhouse, or in a field. Plants may also be progeny of an initial plant comprising a heterologous nucleic acid provided the progeny inherits the heterologous nucleic acid. Seeds produced by a transgenic plant can be grown and then selfed (or outcrossed and selfed) to obtain seeds homozygous for the nucleic acid construct.
- 25
- 30

- The plants to be used with the invention can be grown in suspension culture, or tissue or organ culture. For the purposes of this invention, solid and/or liquid tissue culture techniques can be used. When using solid medium, plant cells can be placed directly onto the medium or can be placed onto a filter that is then placed in contact with the
- 35

medium. When using liquid medium, transgenic plant cells can be placed onto a flotation device, *e.g.*, a porous membrane that contacts the liquid medium.

5 When transiently transformed plant cells are used, a reporter sequence encoding a reporter polypeptide having a reporter activity can be included in the transformation procedure and an assay for reporter activity or expression can be performed at a suitable time after transformation. A suitable time for conducting the assay typically is about 1-21 days after transformation, *e.g.*, about 1-14 days, about 1-7 days, or about 1-3 days. The use of transient assays is particularly convenient for rapid analysis in
10 different species, or to confirm expression of a heterologous polypeptide whose expression has not previously been confirmed in particular recipient cells.

Techniques for introducing nucleic acids into monocotyledonous and dicotyledonous plants are known in the art, and include, without limitation, *Agrobacterium*-mediated
15 transformation, viral vector-mediated transformation, electroporation and particle gun transformation, U.S. Patent Nos 5,538,880; 5,204,253; 6,329,571; and 6,013,863. If a cell or cultured tissue is used as the recipient tissue for transformation, plants can be regenerated from transformed cultures if desired, by techniques known to those skilled in the art.

20

The plant comprising a heterologous nucleic acid to be used with the present invention may for example be selected from: corn (*Zea. mays*), canola (*Brassica napus*, *Brassica rapa* ssp.), alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), sunflower (*Helianthus annuus*), wheat (*Triticum aestivum* and other species), Triticale, Rye (*Secale*) soybean (*Glycine max*), tobacco
25 (*Nicotiana tabacum* or *Nicotiana glauca*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium hirsutum*), sweet potato (*Ipomoea batatas*), cassava (*Manihot esculenta*), coffee (*Coffea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus (*Citrus* spp.) cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea americana*), fig (*Ficus carica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), apple (*Malus* spp.), Pear (*Pyrus* spp.), plum and cherry
30 tree (*Prunus* spp.), *Ribes* (currant etc.), *Vitis*, Jerusalem artichoke (*Helianthemum* spp),

non-cereal grasses (Grass family), sugar and fodder beets (*Beta vulgaris*), chicory, oats, barley, vegetables, and ornamentals.

5 For example, plants of the present invention are crop plants (for example, cereals and pulses, maize, wheat, potatoes, tapioca, rice, sorghum, millet, cassava, barley, pea, sugar beets, sugar cane, soybean, oilseed rape, sunflower and other root, tuber or seed crops. Other important plants maybe fruit trees, crop trees, forest trees or plants grown for their use as spices or pharmaceutical products (*Mentha* spp, clove, *Artemisia* spp, *Thymus* spp, *Lavendula* spp, *Allium* spp., *Hypericum*, *Catharanthus* spp, *Vinca* spp, *Papaver* spp., *Digitalis* spp, *Rawolfia* spp., *Vanilla* spp., *Petrusilium* spp., *Eucalyptus*, tea tree, *Picea* spp, *Pinus* spp, *Abies* spp, *Juniperus* spp., Horticultural plants which may be used with the present invention may include lettuce, endive, and vegetable brassicas including cabbage, broccoli, and cauliflower, carrots, and carnations and geraniums.

15 The plant may also be selected from the group consisting of tobacco, cucurbits, carrot, strawberry, sunflower, tomato, pepper and *Chrysanthemum*.

The plant may also be a grain plants for example oil-seed plants or leguminous plants. 20 Seeds of interest include grain seeds, such as corn, wheat, barley, sorghum, rye, etc. Oil-seed plants include cotton soybean, safflower, sunflower, Brassica, maize, alfalfa, palm, coconut, etc. Leguminous plants include beans and peas. Beans include guar, locust bean, fenugreek, soybean, garden beans, cowpea, mung bean, lima bean, fava bean, lentils, chickpea.

25 In a further embodiment of the invention said plant is selected from the following group: maize, rice, wheat, sugar beet, sugar cane, tobacco, oil seed rape, potato and soybean. Thus, the plant may for example be rice.

30 The whole genome of *Arabidopsis thaliana* plant has been sequenced (The Arabidopsis Genome Initiative (2000). "Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*". *Nature* **408** (6814): 796–815. [doi:10.1038/35048692](https://doi.org/10.1038/35048692). PMID 11130711). Consequently, very detailed knowledge is available for this plant and it may therefore be a useful plant to work with.

35

Accordingly, one plant, which may be used with the present invention is an *Arabidopsis* and in particular an *Arabidopsis thaliana*.

5 In one embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- a) a heterologous nucleic acid encoding Ossyn-CPP of SEQ ID NO:1 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and
- 10 b) a heterologous nucleic acid encoding SsSCS of SEQ ID NO:11 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formulas XXVI and/or XXVII, for example for production of compound 11
15 shown in figure 2.

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- 20 a) a heterologous nucleic acid encoding Ossyn-CPP of SEQ ID NO:1 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and
- b) a heterologous nucleic acid encoding MvTPS5 of SEQ ID NO:18 or a functional
25 homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formulas II, VI, XXXVIII, XXXV, or XXXVI, for example for production of
30 compounds 6, 19 and/or 22 shown in figure 2B.

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- 35 a) a heterologous nucleic acid encoding Ossyn-CPP of SEQ ID NO:1 or a functional homologue thereof sharing at least 70%, such as at least 80%, such

as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and

- b) a heterologous nucleic acid encoding CfTPS4 of SEQ ID NO:13 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formulas II, VI, XXXVIII, XXXV, or XXXVI, for example for production of compounds 6, 19 and/or 22 shown in figure 2B.

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- a) a heterologous nucleic acid encoding Ossyn-CPP of SEQ ID NO:1 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and
- b) a heterologous nucleic acid encoding CfTP3 of SEQ ID NO:12 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formulas II, VI, XXXVIII, XXXV, or XXXVI, for example for production of compounds 6, 19 and/or 22 shown in figure 2B.

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- a) a heterologous nucleic acid encoding EpTPS7 of SEQ ID NO:2, ZmAN2 of SEQ ID NO:3 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and
- b) a heterologous nucleic acid encoding SsSCS of SEQ ID NO:11 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least

85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formulas XXVI or XXVIII, for example for production of compound 23b shown in figure 2B.

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- 10 a) a heterologous nucleic acid encoding EpTPS7 of SEQ ID NO:2, ZmAN2 of SEQ ID NO:3 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and
- 15 b) a heterologous nucleic acid encoding TwTPS2 of SEQ ID NO:14 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formulas IV or X, for example for production of compounds 15, 21 or 45 shown in figure 2B.

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- 25 a) a heterologous nucleic acid encoding EpTPS7 of SEQ ID NO:2, ZmAN2 of SEQ ID NO:3 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and
- 30 b) a heterologous nucleic acid encoding EpTPS1 of SEQ ID NO:15 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formula X, for example for production of compound 21 shown in figure 2B.

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- 5 a) a heterologous nucleic acid encoding EpTPS7 of SEQ ID NO:2, ZmAN2 of SEQ ID NO:3 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and
- 10 b) a heterologous nucleic acid encoding CfTPS14 of SEQ ID NO:16 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formula X, for example for production of compound 21 shown in figure 2B.

15

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- 20 a) a heterologous nucleic acid encoding EpTPS7 of SEQ ID NO:2, ZmAN2 of SEQ ID NO:3 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and
- 25 b) a heterologous nucleic acid encoding EpTPS8 of SEQ ID NO:9 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formulas I, II, VI, XXII, XXIII or XXIV, for example for production of compounds 22, 27a/b or 34 shown in figure 2B.

30

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- 35 a) a heterologous nucleic acid encoding EpTPS7 of SEQ ID NO:2, ZmAN2 of SEQ ID NO:3 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%,

such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and

- b) a heterologous nucleic acid encoding EpTPS23 of SEQ ID NO:10 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formula II or XXIV, for example for production of compound 9a/b shown in figure 2B.

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- a) a heterologous nucleic acid encoding TwTPS7 of SEQ ID NO:4, CfTPS1 of SEQ ID NO:5 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and
- b) a heterologous nucleic acid encoding EpTPS8 of SEQ ID NO:9 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formula I, II, XXIII or XXIV, for example for production of compounds 9a/b or 27a/b shown in figure 2B.

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- a) a heterologous nucleic acid encoding TwTPS7 of SEQ ID NO:4, CfTPS1 of SEQ ID NO:5 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and
- b) a heterologous nucleic acid encoding CfTPS4 of SEQ ID NO:13 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least

85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formulas VI, XXXIX or XL, for example for production of compounds 22 or 25 shown in figure 2B.

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- a) a heterologous nucleic acid encoding TwTPS7 of SEQ ID NO:4, CfTPS1 of SEQ ID NO:5 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and
- b) a heterologous nucleic acid encoding CfTPS3 of SEQ ID NO:12 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formulas VI, XXXIX or XL, for example for production of compounds 22 or 25 shown in figure 2B.

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- a) a heterologous nucleic acid encoding TwTPS7 of SEQ ID NO:4, CfTPS1 of SEQ ID NO:5 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and
- b) a heterologous nucleic acid encoding MvTPS5 of SEQ ID NO:18 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formulas VI, XXXIX or XL, for example for production of compounds 22 or 25 shown in figure 2B.

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- 5 a) a heterologous nucleic acid encoding TwTPS7 of SEQ ID NO:4, CfTPS1 of SEQ ID NO:5 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and
- 10 b) a heterologous nucleic acid encoding SsSCS of SEQ ID NO:11 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formulas XXVI or XXIX, for example for production of compound 23a shown in figure 2B.

15

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- 20 a) a heterologous nucleic acid encoding SsLPPS of SEQ ID NO:6, CfTPS2 of SEQ ID NO:17 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and
- 25 b) a heterologous nucleic acid encoding MvTPS5 of SEQ ID NO:18 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formulas III or XXV, for example for production of compound 16a shown in figure 2B.

30

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- 35 a) a heterologous nucleic acid encoding SsLPPS of SEQ ID NO:6, CfTPS2 of SEQ ID NO:17 or a functional homologue of any of the aforementioned sharing

at least 70%, such as at least 80%, such as at least 85%, such as at least 90%,
such as at least 95%, such as at least 98%, such as at least 99% sequence
identity therewith; and

- 5 b) a heterologous nucleic acid encoding SsSCS of SEQ ID NO:11 or a functional
homologue thereof sharing at least 70%, such as at least 80%, such as at least
85%, such as at least 90%, such as at least 95%, such as at least 98%, such as
at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a
core of formulas III, XXV, XXVI, XXX, XXXI, XXXII, XXXIII or XXXIV for example for
10 production of compounds 3, 16a, 16b, 20, 23a/b, 26, 30, 36 or 43 shown in figure
2B.

In another embodiment of the invention, the host organism may comprise at least the
following heterologous nucleic acids:

- 15 a) a heterologous nucleic acid encoding TwTPS21 of SEQ ID NO:7 or a functional
homologue thereof sharing at least 70%, such as at least 80%, such as at least
85%, such as at least 90%, such as at least 95%, such as at least 98%, such as
at least 99% sequence identity therewith; and
- 20 b) a heterologous nucleic acid encoding SsSCS of SEQ ID NO:11 or a functional
homologue thereof sharing at least 70%, such as at least 80%, such as at least
85%, such as at least 90%, such as at least 95%, such as at least 98%, such as
at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a
core of formulas III, XXV, XXVI, XXX, XXXI, XXXII, XXXIII or XXXIV for example for
25 production of compounds 3, 16a, 16b, 20, 23a/b, 26, 30, 36 or 43 shown in figure
2B.

In another embodiment of the invention, the host organism may comprise at least the
following heterologous nucleic acids:

- 30 a) a heterologous nucleic acid encoding TwTPS21 of SEQ ID NO:7 or a functional
homologue thereof sharing at least 70%, such as at least 80%, such as at least
85%, such as at least 90%, such as at least 95%, such as at least 98%, such as
at least 99% sequence identity therewith; and
- 35 b) a heterologous nucleic acid encoding CfTPS3 of SEQ ID NO:12 or a functional
homologue thereof sharing at least 70%, such as at least 80%, such as at least

85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formulas III or XXXII for example for production of compound 16b shown in figure 2B.

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

a) a heterologous nucleic acid encoding TwTPS21 of SEQ ID NO:7 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and

b) a heterologous nucleic acid encoding TwTPS2 of SEQ ID NO:14 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formulas III or XXXII for example for production of compound 20 shown in figure 2B.

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

a) a heterologous nucleic acid encoding TwTPS21 of SEQ ID NO:7 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and

b) a heterologous nucleic acid encoding CfTPS14 of SEQ ID NO:16 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formulas III or XXXII for example for production of compound 20 shown in figure 2B.

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- 5 a) a heterologous nucleic acid encoding TwTPS21 of SEQ ID NO:7 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and
- 10 b) a heterologous nucleic acid encoding EpTPS1 of SEQ ID NO:15 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formulas III or XXXII for example for production of compound 20 shown in figure 2B.

15 In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- 20 a) a heterologous nucleic acid encoding TwTPS14/28 of SEQ ID NO:8 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and
- b) a heterologous nucleic acid encoding SsSCS of SEQ ID NO:11 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

25 Such a host organism is in particular useful for production of diterpenes having a core of formula XXXIII, for example for production of compound 26 shown in figure 2B.

30 In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- 35 a) a heterologous nucleic acid encoding TwTPS14/28 of SEQ ID NO:8 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and

- 5 b) a heterologous nucleic acid encoding MvTPS5 of SEQ ID NO:18, CfTPS3 of SEQ ID NO:12, CfTPS4 of SEQ ID NO:13 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:

- 10 a) a heterologous nucleic acid encoding MvTPS1 of SEQ ID NO:28 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and
- 15 b) a heterologous nucleic acid encoding SsSCS of SEQ ID NO:11 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formula XLI, for example for production of compound 5 shown in figure 2B.

- 20 In another embodiment of the invention, the host organism may comprise at least the following heterologous nucleic acids:
- 25 a) a heterologous nucleic acid encoding MvTPS1 of SEQ ID NO:28 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith; and
- 30 b) a heterologous nucleic acid encoding CfTPS3 of SEQ ID NO:12, CfTPS4 of SEQ ID NO:13, EpTPS8 of SEQ ID NO:9, EpTPS23 of SEQ ID NO:10 or a functional homologue of any of the aforementioned sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

Such a host organism is in particular useful for production of diterpenes having a core of formula XLI, for example for production of compound 5 shown in figure 2B.

- 35 It may be preferred that the host organism does not naturally produce the diterpene to be produced by the methods of the invention.

Sequences

SEQ ID NO:1- Os syn-CPP

5 MPVFTASFQCVTLFGQPASAADAQPLLQGQRPFLHLHARRRRPCGPMISKSPYPASEETREWEAEGQH
EHTDELRETTTTMIDGIRTALRSIGEGEISISAYDTSLVALLKRLDGGDGPQFPSTIDWIVQNQLPDGSWGDA
SFFMMGDRIMSTLACVVALKSWNIHTDKCERGLLIQENMWRLAHEEEDWMLVGFEIALPSLLDMAKDL
DLDIPYDEPALKAIYAERERKLAKIPRDVLHAMPTTLLHSLEGMVLDLWEKLLKLRLDGSFHCSPASTATAF
10 QQTGDQKCFEYLDGIVKKFNNGGVPCYPLDVYERLWAVDRLTRLGISRHFTSEIEDCLDYIFRNWTPDGLAHT
KNCPVKDIDDTAMGFRLLRLYGYQVDPVLLKKFEKDGGKFFCLHGESPSSVTPMYNTYRASQLKFPGGDDGV
LGRAEVFCRSFLQDRRGSNRMKDKWAIKADIPGEVEYAMDYPWKASLPRIETRLYLDQYGGSGDVWIGKV
LHRMTLFCNDLYLKAADFSNFQKECRVELNGLRRWYLRNLERFGGTDPTTLMTSYFLASANIFEPNRA
AERLGWARVALLADAVSSHFRIGGPKNLTSNLEELISLVPFDDAYSGLREAWKQWLMAWTAKESSQESI
EGDTAILLVRAIEFGGRHVLTGQRPDLWEYSQLEQLTSSICRKYRRVLAQENGKSTEKVEEIDQQLDLEMQ
ELTRRVLQGCSAINRLTRETFLHVVKSFYVAYCSPETIDNHIDKVIFQDVI*

15

SEQ ID NO:2 - EpTPS7

MAAAANPSNSILNHLLSSAAARSVSTSQLLFHSRPLVLSGAKDKRDSFVFRIKCSAVSN
PRIQEQTDFVQKNGLPVIKWHFVETDIDHEQVSKSVSNEIKKRVESIKAILESMEDGD
ITISAYDTAWVALVEDINGSGAPQFPASLQWIANNQLPDGSWGDAEIFTAHDRIINTLSC
20 VVALKSWNIHPDMCERGMKYFRENCKLEDENIEHMPIGFEVAFPSLLELAKKLEIQVPE
DSPVLKDVYDSRNLLKKIPKDIMHKVPTTLLHSLEGMPGLEWEKLLKLQSKDGSFLFSP
SSTAYALMQTKDQNCLEYLTKIVHKFNNGGVPNVYPVDLFEHIWAVDRLQRLGISRYFQPPQ
LKDSVDYVARYWEEDGICWARNSSVHDVDDTAMGFRVLRSGHHVSADVFHFKKGDFTF
CFAGQSTQAVTGMYNLLRASQLMFPGEKILEEAKQFSSAFLVKQDANEVLDKWIITKDL
25 PGEVKYALDIPWYASLPRVESRFYIEQYGGSDDVWIGKTLYRMPIVNNDYELKLAKLDYN
NCQAVHRSEWDNIQKWYEESDLAIEFGVSRREILMAYYLAASIFEPEKSRERIAWAKTSV
LLNTIQAYFHENNSTIHEKAQFVQLFKSGFAINARKLEGKTMKLGRIIVGTLNDVSLDT
AMAYGKDISRDLRHAWDICLQKWEESGDMHQGEAQLIVNTINLTSDAWNFDNLSSHQHF
FQLVNEICYKLRKYKKNKVNKKKTTTPEIESHMQLVKLVLESSDDLSDNLKQIFLTVA
30 RSFYYPVAVCDAGTINYHIARVLFERVY*

30

SEQ ID NO:3 - ZmAN2

MVLSSSCTTVPHLSSLAVVQLGPWSSRIKKKTDVAVPAAAGRWRRALARAQHTSESAAV
AKGSSLTPIVRTDAESRRTRWPTDDDDAEPLVDEIRAMLTSMDSGDISVSAYDTAWVGLV
35 PRLDGGEGPQFPAAVRWIRNNQLPDGSWGDAALFSAYDRLINTLACVVTLTRWSLEPEMR
GRGLSFLGRNMWKLATEDEESMPIGFELAFPSLIELAKSLGVHDFPYDHQALQGIYSRE
IKMKRIPKEVMHTVPTSILHSLEGMPGLDWAKLLKLQSSDGSFLFSPAATAYALMNTGDD
RCFSYIDRTVKKFNNGGVPNVYPVDLFEHIWAVDRLRLGISRYFQKEIEQCMDYVNRHWT
EDGICWARNSDVKEVDDTAMAFRLRLHGSVSPDVFKNFEKDGEFFAFVGQSNQAVTGM
40 YNLNRASQISFPGEDVLHRAGAFSYEFLLRKEAEGALRDKWIISKDLPGEVVYTLDFPWY
GNLPRVEARDYLEQYGGGDDVWIGKTLYRMPVNDVYELARMDFNHCQALHQLQLEWQGL
KRWYTENRLMDFGVAQEDALRAYFLAAASVYEPCAAERLAWARAAAILANAVSTHLRNSP
SFRERLEHSLRCRPSEETDGSWFNSSSGSDAVLVKAVLRLTDSLAREAQPIHGGDPEDII
HKLLRSAWAWEVREKADAADSVCNGSSAVEQEGSRMVHDKQTCLLARMIEISAGRAAGE
45 AASEGDRIIQLTGSICDSLKQKMLVSQDPEKNEEMMSHVDELKLRIFEVQYLLRLG
EKKTGSSETRQTFSLIVKSCYAAHCPPHVDRHISRIVIFPVSAAK*

45

SEQ ID NO:4 - TwTPS7

MHSLLMKKVIMYSSQTTHVFPSPHCTIPKSSFFLDAPVVRHLCLSGHGAKKKRLHFDI

QQGRNAISKTHTPEDLYAKQEYSVPEIVKDDDKEEEVVKIHEVVDIISMLSSMEDGEIS
ISAYDTAWVALIQDIHNNGAPQFPSSLLWIAENQLPDGSWGDSRVFLAFDRIINTLACVV
ALKSWNVHPDKCERGIFLKENISMLEKDDSEHMLVGFEFGFPVLLDMARRLGIDVPDDS
PFLQEYVQRDLKLKRIKPDILHNAPTTLLHSLEAIPDLWDTKLLKLCQDGSLLFSPSS
5 TAMAFINTKDENCLRYLNYVVQRFNGGAPTVPYDLFEHNWAVDRLQRLGISRFFQPEIR
ECMSYVYRYWTKDGIFCTRNSRVHDVDDTAMGFRLLRLHGYEVHPDAFRQFKKGCEFICY
EGQSHPTVTVMYNLYRASQLMFPEEKILDEAKQFTEKFLGEKRSANKLLDKWIITKDLPG
EVGFALDVPWYASLPRVEARFFIQHYGGEDDVWLDKALYRMPYVNNNVYLEAKLDYNYC
QALHRTEWGHQKWYEECKPRDFGISRECLLRAYFMAAASIFEPERSMERLAWAKTAILL
10 EIIVSYFNEVGNSTEQRIAFTEFSIRASPMGGYINGRKLKIGTTQELIQMLLATIDQF
SQDAFAAYGHDITRHLHNSWKMWLLKWQEEGDRWLGEAELLIQTINLMADHKIAEKLFMG
HTNYEQLFSLTNKVCYSLGHHELQNNKELEHDMQRLVQLVLTNSSDGDSDIKKTFLLAVA
KRFYYTAFVDPETVNVHIAKVLFERVD*

15 SEQ ID NO:5 - CfTPS1
MGSLSTMNLNHSPMSYSGILPSSSAKAKLLLPGCFSISAWMNNNGKNLNCQLTHKKISKVA
EIRVATVNAPPVHDQDDSTENQCHDAVNIEDPIEYIRTLRTTGDGRISVSPYDTAWVA
LIKDLQGRDAPEFPSSLEWIIQNQLADGSWGDAKFFCVYDRLVNTIACVVALRSWDVHAE
KVERGVRYINENVEKLRDGNEEHMTCGFEVVPALLQRAKSLGIQDLPYDAPVIQEIYHS
20 REQSKRIPLEMMHKVPTSLLFSLEGLNLEWDKLLKLSADGSFLTSPSSTAFAFMQTR
DPKCYQFIKNTIQTFFNGGAPHTYPVDVFGRLWAIDRLQRLGISRFFESEIADCIAHIHRF
WTEKGVFSGRESEFCDDTSMGVRMLMRMHGYDVPDNLKFNFKDDKFSCYGGQMIESPS
PIYNLYRASQLRFPGEQILEDANKFAYDFLQEKLAHNQILDKWVISKHLPDEIKLGLEMP
WYATLPRVEARYYIQQYAGSGDVWIGKTLYRMPISNDTYHELAKTDFKRCQAQHQFEWI
25 YMQEWYESCNEEFGISRKELLVAYFLATASIFELERANERIAWAKSQIISTIIASFFNN
QNTSPEDKLAFLTDKNGNSTNMALVTLTQFLEGFDRTSHQLKNAWSVWLRLKQQGEGN
GGADAELLVNTLNICAGHIAFREEILAHNDYKTLNLSKICRQLSQIQNEKELETEGQK
TSIKNKELEEDMQRLVKLVLEKSRVGINRDMKKTFLAVVKTYYYKAYHSAQAIDNHMFKV
LFEPVA*

30 SEQ ID NO:6 - SsLPPS
MTSVNLSRAPAAITRRRLQLQPEFHAECSWLKSSSKHAPLTLSCQIRPKQLSQIAELRVT
SLDASQASEKDISLVQTPHKVEVNEKIEESIEYVQNLLMTSGDGRISVSPYDTAVIALIK
DLKGRDAPQFPSCLEWIAHHQLADGSWGDEFFCIYDRILNTLACVVALKSWNLHSDIIEK
35 GVTYIKENVHKLKGANVEHRTAGFELVPTFMQMATDLGIQDLPYDHPHPLIKEIADTKQQR
LKEIPKDLVYQMPTNLLYSLEGLDLEWERLLKLSGNGSFLTSPSSTAANLHMTKDEKC
LKYIENALKNCDGGAPHTYPVDIFSRLWAIDRLQRLGISRFFQHEIKYFLDHIESVWEET
GVFSGRYTKFSDIDDTSMGVRLLKMHGYDVPDNLKHFQKQDGKFCYIGQSVESASPMY
NLYRAAQLRFPGEVLEEATKFAFNFLQEMLVKDRLQERWVISDHLFDEIKLGLKMPWYA
40 TLPRVEAAYYLDHYAGSGDVWIGKSFYRMPEISNDTYKELAILDFNRCQTQHQLWEIHMQ
EWYDRCSLSEFGISKRELLRSYFLAAATIFEPERTQERLLWAKTRILSKMITSFVNISGT
TSLDYNFNGLDEIISSANEDQGLAGTLLATFHQLLDGFDIYTLHQLKHVWSQWFMKVQQ
GEGSGGEDAVLLANTLNICAGLNEDVLSNNEYTALSTLTNKCINLAQIQDNKILQVVDG
SIKDKLEEQDMQALVKLVLEKSGAVDRNIRHTFLSVSKTFYYDAYHDETTDLHIFKVL
45 FRPVV*

SEQ ID NO:7 - TwTPS21
MFMSSSSSSHARRPQLSSFSYLHPPLPFPGLSFFNTRDKRVNFDSTRIICIAKSKPARTT
PEYSDVLQTLPLIVEDDIQEQUEEPLEVSLENQIRQGVDIVKSMLGSMEDGETSISAYDT
50 AWVALVENIHHPGSPQFPSSLQWIANNQLPDGSWGDPDVLFLAHDRLINTLACVIALKKWN

IHPHKCKRGLSFVKENISKLEKENEHMLIGFEIAFPSLLEMAKKLGIEIPDDSPALQDI
 YTKRDLKLTRIPKDKMHNVPPTLLHSLEGLPDLDWEKLVKLQFQNGSFLFSPSSTAFAM
 HTKDGNCLSYLNLDLVHKFNNGVPTAYPVDLFEHIWSVDRLQRLGISRFFHPEIKECLGYV
 HRYWTKDGICWARNSRVQDIDDTAMGFRLRLHGYEVSPDVFKQFRKGDEFVCFMGQSNQ
 5 AITGIYNLYRASQMMFPEETILEEAKKFSVNFLEKRAASELLDKWIITKDLNPNEVGFA
 DVPWYACLPRVETRLYIEQYGGQDDVWIGKTLRMPYVNNNVYLELAKLDYNNCQSLHRI
 EWDNIQKWYEGYNLGGFGVNKRSLRTYFLATSNIFEPERSVERLTWAKTAILVQAIASY
 FENSREERIEFANEFQKFPNTRGYINGRRLDVKQATKGLIEMVFATLNQFSLDALVVHGE
 DITHHLYQSWEKWVLTWQEGGDRREGEAELLVQTINLMAGHTHSQEEELYERLFKLTNTV
 10 CHQLGHYHHLNKDKQPQQVEDNGGYNNSNPESISKLQIESDMRELVQLVLNSSDGMDSNI
 KQTF LAVTKSFYYTAFTHPGTVNYHIAKVLFERVV*

SEQ ID NO:8 - TwTPS14/28
 MFMSSSSSSHARRPQLSSFSYLHPPLFPGLSFFNTRDKRVNFDSTRIICIAKSKPARTT
 15 PEYSDVLQTLPLIVEDDIQEEPLEVSLNQIRQGV DIVKSMGLGSMEDGETSISAYDT
 AWVALVENIHPGSPQFPSSLQWIANNQLPDGSGWDPDVFLAHDRLINTLACVIALKKWN
 IHPHKCKRGLSFVKENISKLEKENEHMLIGFEIAFPSLLEMAKKLGIEIPDDSPALQDI
 YTKRDLKLTRIPKDIMHNVPPTLLYSLEGLPSLDWEKLVKLQCTDGSFLFSPSSTACALM
 HTKDGNCFSYINNVLVHKFNNGVPTVYPVDLFEHIWCVDRLQRLGISRFFHPEIKECLGYV
 20 HRYWTKDGICWARNSRVQDIDDTAMGFRLRLHGYEVSPDVFKQFRKGDEFVCFMGQSNQ
 AITGIYNLYRASQMMFPEETILEEAKKFSVNFLEKRAASELLDKWIITKDLNPNEVGFA
 DVPWYACLPRVETRLYIEQYGGQDDVWIGKTLRMPYVNNNVYLELAKLDYNNCQSLHRI
 EWDNIQKWYEGYNLGGFGVNKRSLRTYFLATSNIFEPERSVERLTWAKTAILVQAIASY
 FENSREERIEFANEFQKFPNTRGYINGRRLDVKQATKGLIEMVFATLNQFSLDALVVHGE
 25 DITHHLYQSWEKWVLTWQEGGDRREGEAELLVQTINLMAGHTHSQEEELYERLFKLTNTV
 CHQLGHYHHLNKDKQPQQVEDNGGYNNSNPESISKLQIESDMRELVQLVLNSSDGMDSNI
 KQTF LAVTKSFYYTAFTHPGTVNYHIAKVLFERVV*

SEQ ID NO:9 - EpTPS8
 30 MQVSLSLTTGSEPCITRIHAPSDAPLKQRNNEREKGTLELNGKVSLLKMGEMLRITIENVP
 IVGSTSSYDTAWVGMVPCSSNSSKPLFPESLKWIMENQNPEGNWAVDHAHHPLLLKDSLS
 STLACVLALHKWNLAPQLVHSGLDFIGSNLWAAMDFRQRSPLGFDVIFPGMIHQAIIDLGI
 NLPFNSSSIENMLTNPLLDIQSFEAGKTSHIAYFAEGLGSRLKDWEQLLQYQTSNGSLFN
 SPSTAAAAIHLRDEKCLNYLHSLTKQFDNGAVPTLYPLDARTRISIIDSLEKFGIHSF
 35 IQEMTILLDQIYSFWKEGNEEFKDPGCCATAFRLLRKHGVDVSSDSLAEFEKKEIFYHS
 SAASAHEIDTKSILELFRASQMILQNEPILDRIYDWTISIFLRDQLVKGLIENKSLYEEV
 NFALGHPFANLDRLEARSYIDNYPYDVPLPKTSYRSSNIDNKDLWTIAFQDFNKCQALH
 RVELDYLEKWVKEYKLDLTKWARQKTEYALFTIGAILSEPEYADARISWSQNTVFVTIVD
 DFFDYGGSLDECRNLINLMHKWDDHLTVGFLSEKVEIVFYSMYGTLNDLAAKAEVRQGRG
 40 VRSHLVNLWIWVMENMLKEREWADYNLVPTFYEYVAAGHITIGLGPVLLIALYFMGYPLS
 EDVVQSQEYKGVYLNVSIIARLLNDRVTVKRESAQGKLVSLFVEHGRGAVDEETSMKE
 VERLVESHKRELLRIVQKTEGSVVPQCKDLAWRVSKVLHLLYMDDDGFTCPVKMLNAT
 NAIVNEPLLLTS*

45 SEQ ID NO:10 - EpTPS23
 MLLASSTSSRFFTKWEPSNKTFSGVSRAQLSQRVKNI VVTPDQVKESESSGTSRLKEM
 LKKVEMPISSYDTAWVAMVPSMEHSRNKLPFNSLKWVMENQQPDGSWCFDDSNHPWLIK
 DSLSTLASVLALKKWNVGQQLIDKGLEYIGSNMWAATDMHQYSPIGFNIIFPSMVEHAN
 KLGLSLSLDHSLFQSMRLNRDMETKSLNGRNMAYVAEGLNGSNWKEVMKYQRRNGSILN
 50 SPATTAALHILNDVKCFEYLDLLTKFQHAVPTLYPFDIYARLCILDELEKLGVDRFVE

IEKMILLDIYRCWLEGSEEILEDPTCCAMAFRFLRMNGYVVSVDVLQGFEEEEKLFHVK
DTKSVELELLKASQLKVSEKEGILDRIYSWATSYLKHQLFNASISDKSLQNEVDYVVKHPH
AILRRIENRNYIENYNTKNVSLRKTSTRFVNVDKRSDLLAHSRQDFNKCQIQFKKELAYL
SRWEKKYGLDKLYARQRLEVYFYSIASNLFEPEFSDARLAWTQYAILTTVVDDFFEYAA
5 SMDELVNLTNLIERWDEHGSEEFKSKEVEILFYAIYDLVNEDAEEKAKKYQGRCIKSHLVH
IWIDILKAMLKESEYVRYNIVPTLDEYISNGCTSISFGAILLIPLYFLGKMSEEVVTSKE
YQKLYMHISMLGRLLNDRVTSQKDMAQGLNSVSLRVLHSGTLTEEEAKEEVDKIIIEKH
RRELLRMVVQTEGSVVPKACKKLFWMTSKELHLFYMTEDCFTCPTKLLSAVNSTLKDPLL
MP*

10

SEQ ID NO:11 SsSCS
MSLAFNVGVTPFSGQRVGSRKEKFPVQGFVTTPNRSRLIVNCSLTTIDFMAKMKEN
FKREDDKFPTTTTLRSEDIPSNLCIIDLQRLGVDQFFQYEINTILDNTFRLWQEKHKVI
YGNVTTHAMAFRLLRVKGYEVSSEELAPYGNQEAVSQQTNDLPMIIELYRAANERIYE
15 EERSLEKILAWTTIFLNKQVQDNSIPDKKLHLKLVFYLRYNYKGITIRLGARRNLELYDMT
YYQALKSTNRFSNLCNEDFLVFAKQDFDIHEAQNQKGLQQLQRWYADCRLDTLNF
RDVVIIANYLASLIIGDHAFDYVRLAFAKTSVLVTIMDDFFDCHGSSQECDKIIELVKEWK
ENPDAEYGSEEEILFMALYNTVNELAERARVEQGRSVKEFLVKLWVEILSAFKIELDT
WSNGTQQSFDEYISSSWLSNGSRLTGLLTMQFVGKLSDEMLMSEECTDLARHVCM
20 VGRLLNDVCSSEEREENIAGKSYSILLATEKDGRKVSDEAIAEINEMVEYHWRKVL
QIVYKKESILPRRCKDVFLEMAKGTIFYAYGINDELTPQQSKEDMKSFVF*

SEQ ID NO:12 CftPS3
MSSLAGNLRVIPFSGNRVQTRTGILPVHQTPMITSKSSAAVKCSLTPTDLMGKIKEVFN
25 REVDTSAAAMTTHSTDIPSNLCIIDLQRLGIDQYFQSEIDAVLHDTYRLWQLKKKIDFS
DITTHAMAFRLLRVKGYEVASDELAPYADQERINLQTDVPTVVELYRAAQERLTEEDST
LEKLYVWTS AFLKQQLTDAIPDKKLHKQVEYYLKNYHGILDRMGVRRNLDLYDISHYKS
LKAHRFYNSNEDILAFARQDFNISQAQHQLQQLQRWYADCRLDTLKFGRDVVRIGN
FLTSAMIGDPELSDLRLAFKHIVLVTRIDFFDHGGPKEESYEILELVKEWKEKPAGEY
30 VSEEVEILFTAVYNTVNELAEMAHIEQGRSVKDLLVKLWVEILSVFRIELDTWTNDTALT
LEEYLSQSWVSIGCRICILISMQFQGVKLSDEMILQSEECTDLCRYVSMVDRLLNDVQTFE
KERKENTGNSVSLQAHHKDERVINEEEACIKVKELAEYNRRKLMQIVYKTGTIFPRKCK
DLFLKACRIGCYLYSSGDEFTSPQQMMEDMKSLVYEPLPISPEANNASGEKMSCVSN*

SEQ ID NO:13 - CftPS4
MSITINLRVIAFPGHGVQSRQGIFAVMEFPRNKNTFKSSFAVKCSLTPTDLMGKIKEKL
SEKVDNSVAAMATDSADMPTNLCIVDSLQRLGVEKYFQSEIDTVLDDAYRLWQLKQK
DIFSDITTHAMAFRLLRVKGYDVSSEELAPYADQEGMNLQTIDLAAVIELYRAAQERVA
EEDSTLEKLYVWTSTFLKQQLLAGAIPDQKLHKQVEYYLKNYHGILDRMGVRKGLDLYD
40 AGYYKALKAADRLVDLCNEDLLAFARQDFNINQAQHRKELEQLQRWYADCRLDKLEF
GRDVVRVSNFLTSAILGDPELSEVRLVFAKHIVLVTRIDFFDHGGPREESHKILELIKELW
KEKPAGEYVSKEVEILYTAVYNTVNELAERANVEQGRNVEPFLRTLWVQILSIFKIELDT
WSDDTALTDDYLNNSWVSIGCRICILMSMQFIGMKLPEEMLLSEECVDLCRHVSMV
DRLLNDVQTFEKERKENTGNAVSLLLAAHKGERAFSEEEAIAKAKYLADCNRRSLMQI
45 VYKTGTIFPRKCKDMFLKVCRIKCYLYASGDEFTSPQQMMEDMKSLVYEPLQIHPPAAA*

SEQ ID NO:14 - TwTPS2
MFDKTQLSVSAYDTAWVAMVSSPNRQAPWFPECVNWLLDNQLSDGSWGLPPHHPSLVKD
50 ALSSTLACLLALKRWGLGEQQMTKGLQFIESNFTSINDEEQHTPIGFNIIFPGMIETAID

MNLNPLRSEDINVMLHNRDLELRRNKLEGREAYLAYVSEG MGKLQDWEMVMKYQRKNGS
 LFNSPSTTAAALSHLGNAGCFHYINSLVAKFGNAVPTVYPSDKYALLCMIESLERLGIDR
 HFSKEIRDVLEETRYCWLQGDDEIFSDADTCAMAFRILRVHGVEVSSDPLTQCAEHHFSR
 SFGGHLKDFSTALELFKASQFVIFPEESGLEKQMSWTNQFLKQEF SNGTTRADRF SKYFS
 5 IEVHDTLKFPFHANVERLAHRRNIEHHHVDNTRILKTSYCFSNISNADFLQLAVEDFNRC
 QSIHREELKHLERWVETKLDRLKFARQKMAYCYFSAAGTCFSPELSDARISWAKNSVLT
 TVADDFDFIVGSEELANLVHLLNWDANGSPHYCSEPVEIIFSALRSTICEIGDKALAW
 QGRSVTHHVIEMWLDLLKSALREAEWARNKVVP TFEYVENGYVSMALGPIVLPVYLIG
 PKVSEEVVRSPEFHNLFKLSICGRLINDTRTFKRESEAGKLNSVLLHMIHSGSGTTEEE
 10 AVEKIRGMIADGRRELLRLVLQEKDSVVP RACKDLFWKMVQVLHLYMDGDGDFSPDMML
 NAVNALIREPISL*

SEQ ID NO: 15 - EpTPS1
 15 MSATPNSFFTSPISAKLGHPKSQSVAESNTRIQLDGTREKIKKMFDKVELSVSPYDTAW
 VAMVPSPNSLEAPYFPECSKWIVDNQLNDGSGVYHRDPLLVKDSISSTLACVLALKRWG
 IGEEKVQNGLEFIELNSASLNDLKQYKPVGFDITFPRMLEHAKDFGLNPLDPKYVEAVI
 FSRDLDLKSGCDSTTEGRKAYLAYISEGIGNLQDWNMVMKYQRRNGSIFDSPSATAAASI
 HLHDASCLRYLRALKKFGNAVPTIYPFIYVRLSMVDAIESLGIARHFQEEIKTVLDET
 20 YRYWLQGNEEIFQDCTTCAMAFRILRANGYNVSSEKLNQFTEDHFSNSLGGYLEDMPVL
 ELYKASQLIFPDELFEKQFSWTSQCLKQKISSGLRHTDGINKHITEEVNDVLKFASYAD
 LERLTNWRRIAVYRANETKMLKTSYRCSNIANEHFLELAVEDFNVCQSMHREELKHLGRW
 VVEKRLDKLFARQKLGICYFSSAASLFAPEMSDARISWAKNAVLTTVVDDFFDVGGSEE
 ELINLVQLIERWDVDGSSHFCSEHVEIVFSALHSTICEIGEKAFA YQGRRMTSHVIKIWL
 25 DLLKSMILTETLWSKSKATPTLNEYMTNGNTSFALGPIVLPALFFVGPKLTDEDLKSHELH
 DLFKTMSTCGRLNDWRSYEREESEGLNAVSLHMIYGNNGSVAATEEEATQKIKGLIESE
 RRELMRLVLQEKDSKIPRCKDLFWKMLKVLHMFYKDDGFTSNQMMKTANSLINQPISL
 HER*

30 SEQ ID NO:16 - CfTPS14
 MSLPLSTCVLFVPKGSQFWSSRFYSASLEVG FQRATSAQIAPLSKSFEETKGRIAKLF
 HKDELSISTYDTAWVAMVPSPTSSEPCFPACLNWLLNQCLDGSWARPHHHPMLKKDVL
 SSTLACILALKKWGVGEEQINRGLHFIELNFASATEKCQITPMGFDIVFPAMLDRARALS
 LNIRLEPTTLNDLMNKRDLNRCYQSSTEREVYRAYIAEGMGKLQNWESVMKYQRKNG
 35 TLFNCPSTTAAFTALRNSDCLNYLHLALNKF GDAVPAVFPLDIYSQLCIVDNLERVGIS
 RHFLTEIQSVLDGTYRSWLQGDQIFMDASTCALAFRTLRMNGYNVSSDPITKLIQEGSF
 SRNTMDINTTLELYRASELILYPDERDLEEHNLR LKTILDQELSGGGFILSRQLGRNINA
 EVKQALESPFYAIMDRMAKRRSIEHYHIDNTRILKTSYCSPNFGNEDFLSLSVEDFNRCQ
 VIHREELRELERWVIENRLDELKFARSKSAYCYFSA AATIFSPELSDARMSWAKNGVLTT
 40 VDDFFDVGGSV EELKNLIQLVELWDVDVSRECISPSVQIIFSALKHTIREIGDKGFKLQ
 GRSITDHIIAIWLDLLYSMMKESEWGREKAVPTIDEYISNAYVSFALGPIVLPALYLVGP
 KLSEEMVNHADYHNLFSMSTCGRLNDIRGYERELKD GKLNLSLYMVNNEGEISWEAA
 ILEVKSWIERERRELLRSVLEEEKSVVPKACKELFWHMCTV VHLFYSKDDGFTSQDLLSA
 VNIIYQPLVLE*

45 SEQ ID NO:17 - CfTPS2
 MKMLMIKSQFRVHSIVSAWANNNSNKRQSLGHQIRRKQRSQVTECRVASLDALNGIQKVG P
 ATIGTPEEENKKIEDSIEYVKELLKTMGDGRISVSPYDTAIVALIKDLEGGDGPEFPSC L
 EWIAQNQLADGSGWDHFFCIYDRVVNTAACVVALKSWNVHADKIEKGAVYLKENVHKLKD
 50 GKIEHMPAGFEFVVPATLERAKALGIGLPYDDPFIREIYSAKQTRLT KIPKGMIIYESPT

5 SLLYSLDGLELEWDKILKLQSDGFSITSVSSTAFVMHTNDLKCHAFIKNALTNCNGG
VPHTYPVDIFARLWAVDRLQRLGISRFFEPEIKYLM DHINNVWREKGVFSSRHSQFADID
DTSMGIRLLKMHGYNVNPNALEHFKQKDGKFTCYADQHIESPSPMYNLYRAAQLRFPGEE
ILQQALQFAYNFLHENLASNHFEKQWVISDHLIDEVRIGLKMPWYATLPRVEASYLQHY
GGSSDVWIGKTLYRMPEISNDTYKILAQLDFNKCQAQHQLWMSMKEWYQSNNVKEFGIS
KKELLAYFLAAATMFEPERTQERIMWAKTQVVSRLMITSFLNKENTMSFDLKIALLTQPP
HQINGSEMKNGLAQLPAAFRQLLEFDKYTRHQLRNTWNKWLMLKQGGDDNGGADAELL
ANTLNICAGHNEDILSHYEYALSSLTNKCQRLSQIQDKKMLEIEEGSIKDKEMELEIQ
10 TLVKLVLQETSGGIDRNIQTFLSVFKTFYYRAYHDAKTIDAHIFQVLFEPV*

SEQ ID NO:18 - MvTPS5
MSITFNLKIAPFSGPGIQRSKETFPATEIQITASTKSTMTTKCSFNASTDFMGKLREKVG
GKADKPPVVIHPVDISSNLCMIDTLQSLGVDYFQSEINTLLEHTYRLWKEKKKNIFKD
VSCCAIAFRLLREKGYQVSSDKLAPFADYRIRDVATILELYRASQARLYEDEHTLEKLH
15 DWSSNLLKQHLLNGSIPDHKLHKQVEYFLKNYHGILDRVAVRRSLDLYNINHHHRIPDV
ADGFPKEDFLEYSMQDFNICQAQQQEELHQLQRWYADCRLDTLNYGRDVVRIANFLT
SAIFGEPEFSDARLAFKHIILVTRIDDFDHGGSREESYKILDLVQEWKEKPAEEYGS
KEVEILFTAVYNTVNDLAEKAHIEQGRCVKPLLIKLVWEILTSFKKELDSWTEETALTLD
EYLSSSWVSGCRICILNSLQYLGIKLSEEMLSSECTDLCRHVSSVDRLNDVQTFKK
20 ERLNTINSVGLQLAAHKGERAMTEEDAMSKIKEMADYHRRKLMQIVYKEGTVFPRE
CKDVFLRVCRIGYYLYSSGDEFTSPQQMKEDMKSLVYQPVKIHPLAINV*

SEQ ID NO:19 – codon optimized DNA sequence encoding truncated CfTPS1:
ATGGGTTCTTGTCTACCATGAACCTGAACCATCTCCAATGCTCTACTCTGGTATTTGCCATCTTCTTC
25 AGCTAAGGCTAAGTTGTTGTTGCCAGTTGTTTTCTATTTCCGCTTGGATGAACAACGGTAAGAATTTG
AATTGCCAATTGACCCACAAGAAGATCTCTAAGGTTGCCGAAATTAGAGTTGCTACTGTTAATGCTCCA
CCAGTTCATGATCAAGATGACTCTACTGAAAATCAATGCCATGATGCCGTTAACAACATCGAAGATCCA
ATCGAATATATCAGAACCTTGTGAGAACTACCGGTGATGGTAGAATTTCTGTTTCTCCATATGATACTG
CTTGGGTCGCTTTGATTAAGGACTTGCAAGGTAGAGATGCTCCAGAATTTCCATCTTCATTGGAATGGA
30 TCATCCAAAATCAATTGGCTGATGGTTCTTGGGGTATGCTAAGTTTTTTCGTTTTACGATAGATTGGT
CAACACCATTGCTTGTGTTGTTGCTTTGAGATCTTGGGATGTTTCATGCTGAAAAAGTTGAAAGAGGTGT
CAGATATATCAACGAAAACGTCGAAAAGTTGAGAGATGGTAACGAAGAACATATGACCTGTGGTTTCG
AAGTTGTTTTCCAGCTTTGTTGCAAAGAGCTAAGTCTTGGGTATTCAAGATTTGCCATATGATGCCCC
AGTTATCCAAGAAATCTACTCTAGAGAACAAAAGTCCAAGAGAATCCATTGGAAATGATGCATAA
35 GGTCCCAACTAGTTTGTGTTCTCTTGGAAAGTTTGGAAAAGTTGGAATGGGACAAGTTGTTGAAGTT
GCAATCAGCAGATGGTTCCTTTTACTCTCTCCATCTTCTACTGCTTTGCTTTTCATGCAAACTAGAGATC
CAAAGTGCTACCAATTCATCAAGAACACCATTCAAACCTTCAACGGTGGTGCTCCACATACTTATCCAGT
TGATGTTTTTGGTAGATTGTGGGCCATTGACAGATTGCAAAGATTGGGTATTTCCAGATTCTTCGAATCC
GAAATTGCTGACTGCATTGCCCATATTCATAGATTCTGGACTGAAAAGGTGTTTTCTGGTAGAGAA
40 TCTGAATTCTGCGATATCGATGATACCTCTATGGGTGTTAGATTGATGAGAATGCATGGTTACGATGTT
GATCCAAACGTCTTGAAGAATTTCAAGAAGGACGATAAGTTCTCTTGCTACGGTGGTCAAATGATTGAA
TCTCCATCTCCAATCTACAACCTGTACAGAGCTTCCAATTGAGATTTCCAGGTGAACAAATTTTGAAG
ATGCCAACAAAGTTGCCTACGACTTTTACAAGAAAAGTTGGCCATAATCAAATCTTGACAAGTGGG
TTATTTCCAAACATTTGCCAGACGAAATCAAGTTGGGTTTGAAGATGCCATGGTATGCTACTTTGCCAAG
45 AGTTGAAGCCAGATATTACATCCAATATTACGCTGGTTCTGGTGATGTTGGATTGGTAAACCTTGTAT
AGAATGCCAGAAATCTCCAACGATACCTATCATGAATTGGCTAAGACCGATTTCAAGAGATGTCAAGCT
CAACATCAATTTGAATGGATCTACATGCAAGAATGGTACGAATCTTGCAACATGGAAGAATTCGGTATC
TCCAGAAAAGAAATTATTGGTCGCTTACTTCTGGCTACCGCTTCTATTTTGAATTGGAAAGAGCCAACG
AAAGAATTGCTTGGGCTAAGTCTCAAATCATCTCTACTATTATCGCTCCTTCTTCAACAATCAAAACACC
50 TCTCCAGAAGATAAGTTGGCTTTCTTGACTGACTTTAAGAACGGTAAGTCTACCAACATGGCTTTGGTTA

CTTTGACCCAATTCTTAGAAGGTTTCGACAGATACACTTCCCACCAATTGAAAAATGCTTGGTCTGTTTG
GTTGAGAAAAGTTGCAACAAGGTGAAGGTAATGGTGGTGCTGATGCTGAATTATTAGTTAACACCTTGA
ACATTTGCGCCGGTCATATTGCTTTCAGAGAAGAAATTTTGGCTCACAACGATTACAAGACCTTGTCTAA
CTTGACCTCTAAGATCTGCAGACAATTGAGTCAAATCCAAAACGAAAAAGAATTGGAAACCGAAGGTC
5 AAAAGACCTCCATTAAGAACAAGAATTAGAAGAAGATATGCAAAGATTAGTCAAGTTGGTCTTGGA
AAGTCCAGAGTTGGTATCAACAGAGACATGAAGAAAACCTTCTTGGCCGTTGTTAAGACCTACTACTAC
AAAGCTTATCATTCGCTCAAGCCATCGATAACCATATGTTTAAGGTTTTGTTGCAACCAGTCGCCTGA

SEQ ID NO:20 - codon optimized DNA sequence encoding truncated CfTPS3:
10 ATGATCACCTCCAAATCTTCCGCTGCTGTTAAGTGTTCTTTGACTACTCCAACTGA
TTTGATGGGTAAGATCAAAGAAGTTTTCAACAGAGAAGTTGATACCTCTCCAGCTG
CTATGACTACTCATTCTACTGATATTCCATCCAACTTGTGCATCATCGATACCTTGC
AAAGATTGGGTATCGACCAATACTTCCAATCCGAAATTGATGCTGTCTTGCATGAT
ACTTACAGATTGTGGCAATTGAAGAAGAAGGACATCTTCTCTGATATTACCACTCA
15 TGCTATGGCCTTCAGATTATTGAGAGTTAAGGGTTACGAAGTTGCCTCTGATGAAT
TGGCTCCATATGCTGATCAAGAAAGAATCAACTTGCAAACCATGATGTTCCAACC
GTCGTCGAATTATACAGAGCTGCACAAGAAAGATTGACCGAAGAAGATTCTACCTT
GGAAAAGTTGTACGTTTGGACTTCTGCTTTCTTGAAGCAACAATTATTGACCGATG
CCATCCCAGATAAGAAGTTGCATAAGCAAGTCGAATATTACTTGAAGAACTACCAC
20 GGTATCTTGGATAGAATGGGTGTTAGAAGAACTTGGACTTGTACGATATCTCCCA
CTACAAATCTTTGAAGGCTGCTCATAGATTCTACAACCTTGTCTAACGAAGATATTTT
GGCCTTCGCCAGACAAGATTTCAACATTTCTCAAGCCCAACACCAAAAAGAATTGC
AACAATTGCAAAGATGGTACGCCGATTGCAGATTGGATACTTTGAAATTCGGTAGA
GATGTCGTCAGAATCGGTAACCTTTTAACTCTGCTATGATCGGTGATCCAGAATT
25 GTCTGATTTGAGATTGGCTTTTGTCTAAGCACATCGTTTTGGTTACCAGAATCGATG
ATTTCTTCGATCATGGTGGTCCAAAAGAAGAATCCTACGAAATTTTGAATTGGTC
AAAGAATGGAAAGAAAAGCCAGCTGGTGAATACGTTTCTGAAGAAGTCGAAATCTT
ATTCACCGCTGTTTACAACACCGTTAACGAATTGGCTGAAATGGCCCATATTGAAC
AAGGTAGATCTGTTAAGGATTTGTTGGTTAAGTTGTGGGTGCAAAATATTGTCCGTT
30 TTCAGAATCGAATTGGATACCTGGACTAACGATACTGCTTTGACTTTGGAAGAATA
CTTGTCCCAATCCTGGGTTTCTATTGGTTGCAGAATCTGCATTTTGATCTCCATGC
AATTCCAAGGTGTTAAGTTGAGTGACGAAATGTTGCAAAGTGAAGAATGTACCGAT
TTGTGCAGATACGTTTCCATGGTCGATAGATTATTGAACGATGTCCAAACCTTCGA
AAAAGAAAAGAAAAGAAAACACCGGTAACCTCCGTTTCTTTGTTGCAAGCTGCTCACA
35 AAGACGAAAGAGTTATCAACGAAGAAGAAGCCTGCATCAAGGTAAAAGAATTAGC
CGAATACAATAGAAGAAAGTTGATGCAAATCGTCTACAAGACCGGTACTATTTTCC
CAAGAAAATGCAAGGACTTGTTCTTGAAGGCTTGTAAGATTGGTTGCTACTTGTAC
TCTTCTGGTGATGAATCACTTCCCCACAACAATGATGGAAGATATGAAGTCCTT
GGTCTATGAACCATTGCCAATTTCTCCACCTGAAGCTAACAATGCATCTGGTGAAA
40 AAATGTCCTGCGTCAGTAACTGA

SEQ ID NO:21 - codon optimized DNA sequence encoding truncated ZmAN2:
ATGGCCCAACATACTTCTGAATCTGCTGCTGTTGCTAAAGGTTCTTCTTTGACTCCAATCGTTAGAACCG
ATGCTGAATCTAGAAGAACTAGATGGCCAACAGATGATGATGACGCTGAACCATTGGTTGACGAAATT
45 AGAGCTATGTTGACCTCTATGTCCGATGGTGATATTTCTGTTTCTGCTTATGATACTGCTTGGGTTGGTT
TGGTTCCAAGATTGGATGGTGGTGAAGGTCCACAATTTCCAGCTGCTGTTAGATGGATTAGAAAACAATC
AATTGCCAGATGGTTCTTGGGGTGATGCTGCTTTGTTTTCACTTACGATAGATTGATTAACACCTTGGC
TTGTGTTGTTACTTTGACCAGATGGTCTTTGGAACAGAAATGAGAGGTAGAGGTTGTCTTTTTTGGG
TAGAAACATGTGGAAGTTGGCTACCGAAGATGAAGAATCTATGCCAATTGGTTTGAATTGGCTTTCCC
50 ATCCTTGATTGAATTGGCTAAATCTTTGGGTGTTACGATTTCCCATATGATCATCAAGCTTTACAAGGT
ATCTACTCCTCCAGAGAAATCAAATGAAGAGAATCCCAAAGAAGTCATGCATACTGTTCCAACCTCT
ATCTTGCAATCTTTGGAAGGTATGCCAGGTTTGGATTGGGCTAAGTTGTTGAAATTGCAATCCTCTGAT

GGTTCATTCTGTTTTACCCAGCTGCTACTGCTTACGCTTTGATGAATACTGGTGATGATAGATGCTTCTC
CTACATTGATAGAACCGTCAAAAAGTTCAATGGTGGTGTCCAAATGTTTACCCAGTTGACTTGTGAA
CATATCTGGGCTGTTGACAGATTGGAAAGATTGGGTATTTCCAGATACTTCAAAAAGAAATCGAACAA
TGCATGGACTACGTTAACAGACATTGGACTGAAGATGGTATTTGTTGGGCTAGAACTCCGACGTAAA
5 AGAAGTTGACGATACTGCTATGGCCTTCAGATTATTGAGATTGCATGGTACTCTGTTTCCCAGATGTT
TTCAAGAACTTCGAAAAGGATGGTGAATTTCTCGCTTCGTCGGTCAATCTAATCAAGCTGTACTGGTA
TGTACAACCTGAACAGAGCCTCCCAAATTCATTTCCAGGTGAAGATGTTTACACAGAGCTGGTGCTTT
TTCTTACGAATTTCTTGAGAAGAAAAGAAGCCGAAGGTGCTTTGAGAGATAAGTGGATTATTTCCAAGG
ATTTGCCTGGTGAAGTTGTCTACACTTTGGATTTTCATGGTACGGTAATTTGCCAAGAGTTGAAGCTA
10 GAGACTACTTGGAACAATATGGTGGTGGTATGACGTTTGGATAGGTAAAACATTATACAGAATGCCA
TTGGTCAACAACGACGTTTATTTGGAATTGGCCAGAATGGATTTCAACCATTGTCAAGCCTTGCATCAAT
TGGAATGGCAAGGTTTGAAGAAGATGGTACACCGAAAACAGATTGATGGATTTTGGTGTGCTCAAGAA
GATGCATTGAGAGCTTACTTTTTGGCTGCTGCTTCAGTTTATGAACCATGTAGAGCTGCTGAAAGATTA
GCTTGGGCAAGAGCTGCTATTTTGGCTAATGCTGTTTCTACTCACTTGAGAACTCTCCATCTTTCAGAG
15 AAAGATTGGAACACTCTTTGAGATGCAGACCTTCTGAAGAACTGATGGTAGTTGGTTCAATTCCTCTT
CTGGTCTGATGCTGTTTGGTTAAGGCAGTTTGGAGATTGACTGATTCCTTGGCTAGAGAAGCTCAAC
CTATTCACGGTGGTATCCAGAAGATATTATTCACAAGTTGTTAAGATCCGCTGGGCTGAATGGGTGA
GAGAAAAAGCTGATGCTGCAGATTCTGTCTGTAATGGTTCTTCTGCTGTTGAACAAGAAGTTCCAGAA
TGGTTCATGATAAGCAAACCTGTTTGTGTTGGCAAGAATGATTGAAATTTCCGCTGGTAGAGCCGCTG
20 GTGAAGCTGCTTCCGAAGATGGTGACAGAAGAATTATACAATTGACCGTTCCATCTGCGACTCATTGA
AACAAAAATGTTGGTCAGTCAAGACCCAGAAAAGAACGAAGAAATGATGTCCCATGTTGACGACGAA
TTGAAGTTGAGAATCAGAGAATTCGTCCAATACTTGTTGAGATTGGGTGAAAAAAGACTGGTTCCTCT
GAAACCAGACAACTTTCTTGTCTATCGTCAAGTCTTGTTACTACGCTGCTCATTGTCCACCACATGTTGT
TGATAGACATATCTCCAGAGTTATCTTGAACCAAGTTTCTGCTGCTAAATTGGAACATCATCACCATCAC
25 CACTGA

SEQ ID NO:22 - codon optimized DNA sequence encoding truncated EpTPS1:
ATGGCTCAATCCGTTGCTGAATCCAACACCAGAATCAACAATTGGATGGTACTAGAGAAAAGATCAAG
AAGATGTTGACAAGGTCGAATTGTCTGTTTCTCCATATGATACTGCTTGGGTTGCTATGGTTCCATCTC
30 CAAATCTTTGGAAGCTCCATACTTTCCAGAATGCTCTAAATGGATCGTCGACAATCAATTGAATGATGG
TTCTTGGGGTTTCTACCATAGAGATCCATTATTGGTTAAGGACTCCATCTCTTCTACTTTGGCTTGTGTTT
TGGCTTTGAAAAGATGGGGTATTGGTGAAGAAGCAAGTCAACAAAGGTTTGAATTATCGAATTGAAC
TCCGCCTCTTTGAACGATTGAAACAATACAAGCCAGTCGGTTTCGATATTACCTTTCCAAGAATGTTGG
AACACGCTAAGGATTTGCGTTTGAATTTGCCATTGGATCCTAAGTATGTTGAAGCCGTTATCTTCTCCAG
35 AGATTTGGATTTGAAATCCGGTTGTGATTCTACTACCGAAGGTAGAAAAGCTTACTTGGCCTATATTTCC
GAAGGTATCGGTAACCTGCAAGATTGGAATATGGTCATGAAGTACCAAAGAAGAAACGGTTCCATTTT
CGATTCTCCATCTGCTACAGCTGCTGCTTCTATTCACTTGCATGATGCTTCATGTTTGAAGATACTTGAGAT
GCGCCTTGAAGAAATTTGGTAATGCTGTTCCAATCTACCCATTCAACATCTACGTCAGATTGTCTAT
GGTTGATGCCATTGAATCTTTGGGTATTGCCAGACACTTCAAGAAGAAATCAAGACCGTTTTGGACGA
40 AACTTACAGATATTGGTTGCAAGGTAACGAAGAAATCTTCCAAGATTGCACTACTTGTGCTATGGCCTT
CAGAATTTGAGAGCTAATGGTTACAACGTTTCTCCGAAAAGTTGAATCAATTCACCGAAGATCACTTC
TCCAATTCATTGGGTGGTTATTTGGAAGATATGAGACCACTTGGAAATATACAAGGCCTCCCAATTG
ATTTCCAGACGAATTATTCTTAGAAAAGCAATTCTCCTGGACCTCCCAATGTTTGAAGCAAAAAATCT
CTTCCGTTTGAAGACATACCGACGGTATTAACAAACACATTACCGAAGAAGTTAACGACGTTTTGAAGT
45 TCGCTTCTTACGCTGATTTGGAAGATTGACCAATTGGAGAAGAATCGCTGTTTACAGAGCTAACGAAA
CAAAAATGTTGAAAACCTCCTACAGATGCTCCAACATTGCTAACGAACACTTTTTGGAATTGGCCGTCG
AAGATTTCAACGTTTGTCAATCAATGCACAGAGAAGAATTGAAGCACTTGGGTAGATGGGTTGTTGAA
AAGAGATTGGACAAGTTGAAATTCGCCAGACAAAAGTTGGGTTACTGCTACTTTTCTCAGCTGCTTCTT
TGTTTGCTCCAGAAATGTCTGATGCTAGAATTTCTGGGCTAAGAATGCCGTTTTGACTACCGTTGTTGA
50 TGACTTTTTTATGATGTCGGTGGTTCGGAAGAAGAATTGATTAACCTGGTCCAATTGATCGAAAGATGGGA

CGTTGATGGTTCCTCTCATTTCTGTTCTGAACATGTCGAAATCGTTTTCTCTGCCTTGCAATTCTACCATTTG
CGAAATAGGTGAAAAGGCTTTTGCTTATCAAGGTAGAAGAATGACCTCCACGTTATTAAGATTTGGTT
GGACTTGTTGAAGTCCATGTTGACTGAACTTTGTGGTCTAAGTCTAAGGCTACTCCAACCTGAACGA
ATATATGACTAACGGTAACACCTCTTTGCTTTGGGTCCAATAGTTTTGCCAGCTTTGTTTTTTGTTGGTC
5 CAAAGTTGACCGACGAAGATTTGAAGTCTCATGAATTGCACGATTTGTTCAAGACCATGTCTACCTGTG
GTAGATTATTGAACGATTGGAGATCTACGAAAGAGAATCTGAAGAAGGTAAATTGAACGCCGTTTCC
TTGCATATGATCTACGGTAATGGTTCTGTTGCTGCTACTGAAGAAGAAGCTACTCAAAGATTAAGGGT
TTGATCGAATCCGAAAGAAGAGAATTGATGAGATTGGTATTGCAAGAAAAGGACTCTAAGATTCCTAG
ACCATGCAAGGATTTGTTCTGGAAGATGTTGAAGGTCTTGACATGTTCTACTTGAAGGATGATGGTTT
10 CACCTCCAATCAAATGATGAAGACTGCTAACTCCTTGATCAATCAACCTATCTCATTGCACGAAAGAGTT
GAACATCATCATCACCATCACTAA

SEQ ID NO:23 - codon optimized DNA sequence encoding truncated TwTPS21:
ATGGGTATCGCTAAATCCAAGCCAGCTAGAACTACTCCAGAATACTCTGATGTTTTACAACTGGTTTGC
15 CATTGATCGTCGAAGATGATATCCAAGAACAAGAAGAACCATTTGGAAGTTTCTTTGAAAAATCAAATCA
GACAAGGTGTCGACATCGTCAAATCTATGTTGGGTTCTATGGAAGATGGTGAAACCTCTATTTCTGCTT
ATGATACTGCTTGGGTTGCCTTGGTTGAAAACATTCATCATCCAGGTAGTCCACAATCCCATCTTCATT
ACAATGGATCGCCAACAATCAATTGCCAGATGGTTCTTGGGGTGATCCAGATGTTTTTTGGCTCATGA
TAGATTGATTAACACCTTGGCTTGCCTTATTGCTTTGAAGAAGTGGAATATCCATCCACACAAATGCAA
20 GAGAGGTTTGTCTTTCGTCAAAGAAAACATTTCTAAGTTGGAAAAAGAAAACGAAGAACACATGTTGA
TCGGTTTCGAAATTGCCTTCCATCCTTGTGGAAATGGCTAAGAAATTGGGTATCGAAATCCAGATG
ATTCTCCAGCTTTACAAGATATCTACACCAAGAGAGATTTGAAGTTGACCAGAATCCCAAAGGATAAGA
TGCATAACGTTCCAACCTACCTGTTGCATTATTGGAAGGTTTGCCAGATTTGGATTGGGAAAAGTTGG
TTAAGTTGCAATTCAAAACGGTTCCTTTTTGTTCTCTCCATCTTCTACTGCTTTTGCCTTTATGCATACCA
25 AGGATGGTAACTGCTTGTCTACTTGAATGATTTGGTTCACAAGTTCAATGGTGGTGTCCAACCTGCTTA
TCCAGTTGATTTGTTGAACACATCTGGTCCGTTGACAGATTGCAAAGATTGGGTATTTCCAGATTCTTC
CACCCAGAAATCAAAGAATGTTGGGTTACGTTTCATAGATACTGGACTAAGGACGGTATTTGTTGGGCT
AGAAATTCCAGAGTTCAAGATATTGATGATACCGCCATGGGTTTCAGATTATTGAGATTGCATGGTTAC
GAAGTTTCCCAGATGTCTTTAAGCAATTCAGAAAGGGTGATGAATTCGTCTGTTTCATGGGTCAATCC
30 AATCAAGCTATTACCGGTATCTACAACTTGTACAGAGCTTCCCAAATGATGTTCCAGAAAGAAACCATTT
TGGAAGAAGCCAAGAAGTTCTCCGTTAACTTCTTGAGAGAAAAGAGAGCTGCCTCTGAATTATTGGAT
AAGTGGATTATACCAAGGACTTGCCAAATGAAGTTGGTTTTGCTTTGGATGTTCCATGGTATGCTTGT
TGCCAAGAGTTGAAACCAGATTGTACATCGAACAATACGGTGGTCAAGATGATGTTGGATAGGTAAG
ACCTGTATAGAATGCCATACGTCAACAACAACGTCTACTTGAATTTGGCCAAATTGGATTACAACAAC
35 TGCCAATCCTTGACAGAAATTGAATGGGACAATATCCAAAAGTGGTACGAAGGTTACAATTTGGGTGG
TTTTGGTGTCAACAAGAGATCCTTATTGAGAACCTACTTTTTGCCACCTCCAACATTTTGAACCAGAA
AGATCTGTCGAAAGATTGACTTGGGCTAAGACTGCTATTTGGTTCAAGCCATTGCTTCCTACTTCGAAA
ACTCTAGAGAAGAAAGAATCGAATTCGCCAACGAATTTCAAAGTTCCCAAACACTAGAGGTTACATCA
ACGGTAGAAGATTGGATGTTAAGCAAGCTACCAAGGGTTTGATCGAAATGGTTTTGCTACCTTGAATC
40 AATTCTCCTTGGATGCCTTAGTTGTTACGGTGAAGATATTACTCATCACTGTACCAATCCTGGGAAAA
ATGGGTTTTGACTTGGCAAGAAGGTGGTGATAGAAGAGAAGGTGAAGCCGAATTATTAGTCCAAACCA
TTAATTGATGGCCGGTCATACTCATAGTCAAGAAGAAGAATTATACGAAAGATTATTCAAGTTGACTA
ACACCGTCTGCCATCAATTGGGTCAATTATCATCATTTGAACAAGGATAAGCAACCACAACAAGTCGAAG
ATAATGGTGGTTACAACAATTCCAACCCAGAATCCATCTCCAAGTTGCAAATTGAATCCGACATGAGAG
45 AATTGGTCCAATTGGTTTTGAACTCCTCTGATGGTATGGACTTAACATCAAGCAAACCTTTCTGGCTGT
TACCAAGTCTTTCTACTACACTGCTTTTACTCATCCTGGTACTGTCAACTACCATATTGCTAAGGTTTTGT
TCGAAAGAGTCGTCTTAGAACATCATCATCACCATCACTGA

SEQ ID NO:24 - codon optimized DNA sequence encoding truncated SsSCS:
50 ATGTCCTTGGCTTTCAACGTTGGTGTTACTCCATTTCTGGTCAAAGAGTCGGTTCAGAAAAGAAAAGT

TTCCAGTTCAAGGTTTCCCAGTTACTACTCCAAATAGATCCAGATTGATCGTCAACTGTTCTTGACTACC
ATTGATTTCATGGCCAAGATGAAGGAAAACCTTCAAGAGAGAAGATGACAAGTTCCCAACTACTACTACC
TTGAGATCTGAAGATATCCCATCCAACCTTGTGCATTATCGATACCTTGCAAAGATTGGGTGTTGACCAAT
TCTTCCAATACGAAATCAACACCATCTTGGACAACACTTTCAGATTGTGGCAAGAAAAGCACAAGGTTA
5 TCTACGGTAATGTTACTACACATGCTATGGCCTTCAGATTATTGAGAGTTAAGGGTTACGAAGTTTCCTC
CGAAGAATTAGCTCCATACGGTAATCAAGAAGCCGTTTCTCAACAACTAACGACTTGCCAATGATCAT
CGAATTATACAGAGCTGCCAACGAAAGAATCTACGAAGAAGAAAGATCCTTGGAAGAGATTTTGGCTT
GGACCACCATTCTTGAACAAGCAAGTTCAAGACAACCTCCATCCCAGATAAGAAGTTGCATAAGTTGG
TCGAATTCTACTTGAGAACTACAAGGGTATCACCATTAGATTAGGTGCCAGAAGAACTTGGAATTAT
10 ACGACATGACTTACTACCAAGCCTTGAAGTCTACCAACAGATTCTCTAAGTTGTGTAACGAAGATTTCTT
GGTTTTCGCCAAGCAAGATTCGATATTCACGAAGCCCAAAATCAAAGGGTTTACAACAATTACAAAG
ATGGTACGCCGATTGCAGATTGGATACTTTGAATTCGGTAGAGATGTCGTCATTATCGCTAACTATTTG
GCCTCCTTGATTATTGGTGATCATGCCCTTGATTACGTCAGATTGGCTTTTGCTAAGACCTCTGTTTTGGT
TACCATCATGGATGATTTCTTCGATTGCCATGGTCTTCTCAAGAATGCGACAAGATAATCGAATTGGTA
15 AAAGAATGGAAAGAAAACCCAGATGCCGAATACGGTCTGAAGAATTGGAAATTTGTTTCATGGCCTT
GTACAACACCGTTAACGAATTGGCTGAAAGAGCTAGAGTTGAACAAGGTAGATCTGTCAAAGAATTTT
TGGTCAAGTTGTGGGTTGAAATCTTGTCCGCTTCAAGATTGAATTGGATACCTGGTCTAACGGTACTC
AACAATCTTCGACGAATATATCTCCTCCTCTTGGTTGTCTAATGGTTCTAGATTGACTGGTTTGTGACC
ATGCAATTTGTTGGTGTCAAATTGTCCGACGAAATGTTGATGTCAGAAGAATGTACTGATTTGGCTAGA
20 CACGTATGTATGGTCGGTAGATTATTGAACGATGTCTGCTCATCTGAAAGAGAAAGAGAAGAAAACAT
TGCCGGTAAGTCTACTCTATTTTGTGGCTACTGAAAGGACGGTAGAAAGGTTTCTGAAGATGAAGC
TATTGCTGAAATCAACGAAATGGTCAATACCATTGGAGAAAGGCTTGCAAATCGTCTACAAGAAAG
AATCCATCTTGCTAGAAGATGCAAGGACGTTTTTTTGGAAATGGCTAAGGGTACTTTTTACGCCTACG
GTATTAACGATGAATTGACCTCTCCACAACAATCCAAAGAAGATATGAAGTCCTTCGTTTTTAA
25

SEQ ID NO:25 codon optimized DNA sequence encoding truncated TwTPS14:
ATGTTTATGTCCTCCTCCTCATCCTCTCATGCTAGAAGACCACAATTGTCATCTTCTCTTACTTGCATCCA
CCATTGCCATTTCCAGGTTTGTCAATTTTCAACACCAGAGACAAGAGAGTCAACTTCGATTCTACCAGAA
TTATCTGCATTGCCAAATCTAAGCCAGCTAGAACTACTCCAGAATACTCCGATGTTTTACAACTGGTTT
30 GCCATTGATCGTCGAAGATGATATCCAAGAACAAGAAGAACCATTTGGAAGTTTCTTTGGAAAATCAAAT
CAGACAAGGTGTCGACATCGTCAAATCTATGTTGGGTTCTATGGAAGATGGTGAAACCTCTATTTCTGC
TTATGATACTGCTTGGGTTGCCTTGGTTGAAAACATTCATCATCCAGGTAGTCCACAATTCCCATCTTCAT
TACAATGGATCGCCAACAATCAATTGCCAGATGGTCTTGGGGTGATCCAGATGTTTTTTGGCTCATG
ATAGATTGATTAACACCTTGGCTTGCCTTATTGCTTTGAAGAAGTGGAATATCCATCCACACAAATGCA
35 AGAGAGGTTTGTCTTTCTGTCAAAGAAAACATTTCTAAGTTGGAAAAAGAAAACGAAGAACACATGTTG
ATCGGTTTCGAAATTGCCTTCCATCCTTGTTAGAAATGGCTAAGAAGTTGGGTATCGAAATCCCAGAT
GATTCTCCAGCTTTACAAGATATCTACACCAAGAGAGATTTGAAGTTGACCAGAATCCAAAGGATATC
ATGCATAACGTTCCAACTACCTGTTGTAATCTTTGGAAGGTTGCCTTCTTGGATTGGGAAAAGTTGG
TTAAGTTGCAATGTACTGACGGTCTTTTTTGTCTCTCCATCTTCTACTGCTTGTGCTTGTATGCATACA
40 AAAGATGGTAACTGCTTCTCCTACATCAACAACCTTGGTCCATAAGTTTAATGGTGGTGTCCAACCTGTTT
ACCCAGTTGATTTGTTTGAACATATCTGGTGCGTTGACAGATTGCAAAGATTGGGTATTTCCAGATTCTT
CCACCCAGAAATCAAAGAATGTTTGGGTTACGTTCTAGATACTGGACCAAGGATGGTATTTGTTGGGC
TAGAAATTCAGAGTTCAAGATATTGATGATACCGCCATGGGTTTCAGATTATTGAGATTGCATGGTTA
CGAAGTTTCCCAGATGTCTTAAAGCAATTCAGAAAGGGTGATGAATTCGTCTGTTTCATGGGTCAATC
45 CAATCAAGCTATTACCGGTATCTACAACCTTGTACAGAGCTTCCCAATGATGTTCCCAGAAGAAACCAT
TTGGAAGAAGCCAAGAAGTTCTCCGTTAACTTCTTGAGAGAAAAGAGAGCTGCCTCTGAATTATTGGAT
AAGTGGATTATACCAAGGACTTGCCAAATGAAGTTGGTTTTGCTTGGATGTTCCATGGTATGCTTGT
TGCCAAGAGTTGAAACCAGATTGTACATCGAACAATACGGTGGTCAAGATGATGTTTGGATAGGTAAG
ACCTTGTATAGAATGCCATACGTCAACAACAACGTCTACTTGGAAATGGCCAAATGGATTACAACAAC
50 TGCCAATCCTTGACAGAATTGAATGGGACAATATCCAAAAGTGGTACGAAGGTTACAATTTGGGTGG

TTTTGGTGTCAACAAGAGATCCTTATTGAGAACCTACTTTTTGGCCACCTCCAACATTTTTGAACCAGAA
 AGATCTGTGCGAAAGATTGACTTGGGCTAAGACTGCTATTTTGGTTCAAGCCATTGCTTCCTACTTCGAAA
 ACTCTAGAGAAGAAAGAATCGAATTCGCCAACGAATTCAAAAAGTTCCCAAACTAGAGGTTACATCA
 ACGGTAGAAGATTGGATGTTAAGCAAGCTACCAAGGGTTTGATCGAAATGGTTTTCGCTACCTTGAATC
 5 AATTCTCCTTGGATGCATTGGTTGTTACGGTGAAGATATTACTCATCACTTGTACCAATCCTGGGAAAA
 ATGGGTTTTGACTTGGCAAGAAGGTGGTGATAGAAGAGAAGGTGAAGCCGAATTATTAGTCCAAACCA
 TTAAGTTGATGGCCGGTCATACTCATAGTCAAGAAGAAGATTATACGAAAGATTATTCAAGTTGACTA
 ACACCGTCTGCCATCAATTGGGTGATTATCATCATTTGAACAAGGACAAGCAACCACAACAAGTCGAAG
 ATAACGGTGGTTACAACAATTCTAACCAGAATCCATCTCCAAGTTGCAAATCGAATCTGACATGAGAG
 10 AATTGGTCCAATTGGTCTTGAATTCCTCTGATGGTATGGACTTAACATCAAGCAAACCTTTCTTGGCTGT
 TACCAAGTCTTTCTACTACACTGCTTTTACTCATCTGGTACTGTCAACTACCATATTGCTAAGGTTTTGT
 TCGAAAGAGTTGTTAA

SEQ ID NO:28 MvTPS1
 15 MASTPTLNLSITTPFVRTKIPAKISLPACSWLDRSSSRHVELNHKFCRKLELKVAMCRA
 SLDVQQVRDEVYSNAQPHELVDKKIEERVKYVKNLLSTMDDGRINWSAYDTAWISLIK
 DFEGRDPCQFPSTLERIAENQLPDGSGWDKDFDCSYDRIINTLACVVALTTWNVHPEI
 NQKGIRYLKENMRKLEETPTVLMTCAFEVFPALLKKARNLGIHDLPHYDMPPIVKEICKIG
 DEKLARIPKKMMEKETTSMLYAAEGVENLDWERLLKL RTPENG SFLSSPAATV VAFM
 20 HTKDEDCLRYIKYLLNKFNGGAPNVYPVDLWSRLWATDRLQRLGISRYFESEIKDLLS
 YVHSYWTDIGVYCTRDSKYADIDDTSMGFRLLRVQGYNMDANVFKYFQKDDKFVCL
 GGQMNGSATATYNLYRAAQYQFPGEQILEDARKFSQQFLQESIDTNLLDKWVISPHI
 PEEMRFGMEMTWYSCLPRIEASYLQHYGATEDVWLGKTFFRMEEISNENYRELAIL
 DFSKCQAQHQTIEWIHMQEWEYESNNVKEFGISRKDLLFAYFLAAASIFETERAKERILW
 25 ARSKIICKMVKSFLEKETGSLEHKIAFLTSGDGKNGPNNAMATLHQLLGEFDGYISI
 QLENAAWAWLTKEQGEANDGELLATTINICGGRVNQDTLSHNEYKALSDLTNKICHN
 LAQIQNDKGDEIKDSKRSEKDKEVEQDMQALAKLVFEESDLERSIKQTF LAVVRTYYY
 GAYIAAEKIDVHMFKVLFPVG*

SEQ ID NO:1	Amino acid sequence of syn-CPP from <i>Oryza sativa</i>
SEQ ID NO:2	Amino acid sequence of TPS7 from <i>Euphobia peplus</i>
SEQ ID NO:3	Amino acid sequence of AN2 from <i>Zea Maiz</i>
SEQ ID NO:4	Amino acid sequence of TPS7 from <i>Tripterygium Wilfordii</i>
SEQ ID NO:5	Amino acid sequence of TPS1 from <i>Coleus forskohlii</i>
SEQ ID NO:6	Amino acid sequence of LPPS from <i>Salvia scarea</i>
SEQ ID NO:7	Amino acid sequence of TPS21 from <i>Tripterygium Wilfordii</i>
SEQ ID NO:8	Amino acid sequence of TPS14/28 from <i>Tripterygium Wilfordii</i>
SEQ ID NO:9	Amino acid sequence of TPS8 of <i>Euphobia peplus</i>
SEQ ID NO:10	Amino acid sequence of TPS23 of <i>Euphobia peplus</i>
SEQ ID NO:11	Amino acid sequence of SCS of <i>Salvia scarea</i>
SEQ ID NO:12	Amino acid sequence of TPS3 of <i>Coleus forskohlii</i>
SEQ ID NO:13	Amino acid sequence of TPS4 of <i>Coleus forskohlii</i>
SEQ ID NO:14	Amino acid sequence of TPS2 of <i>Tripterygium Wilfordii</i>

SEQ ID NO:15	Amino acid sequence of TPS1 of <i>Euphobia peplus</i>
SEQ ID NO:16	Amino acid sequence of TPS14 of <i>Coleus forskohlii</i>
SEQ ID NO:17	Amino acid sequence of TPS2 of <i>Coleus forskohlii</i>
SEQ ID NO:18	Amino acid sequence of TPS5 from <i>Marrubium vulgare</i>
SEQ ID NO:19	DNA sequence encoding truncated CfTPS1 codon optimised for expression in <i>Saccharomyces cerevisiae</i>
SEQ ID NO:20	DNA sequence encoding truncated CfTPS3 codon optimised for expression in <i>Saccharomyces cerevisiae</i>
SEQ ID NO:21	DNA sequence encoding truncated ZmAN2 codon optimised for expression in <i>Saccharomyces cerevisiae</i>
SEQ ID NO:22	DNA sequence encoding truncated EpTPS1 codon optimised for expression in <i>Saccharomyces cerevisiae</i>
SEQ ID NO:23	DNA sequence encoding truncated TwTPS21 codon optimised for expression in <i>Saccharomyces cerevisiae</i>
SEQ ID NO:24	DNA sequence encoding truncated SsSCS codon optimised for expression in <i>Saccharomyces cerevisiae</i>
SEQ ID NO:25	DNA sequence encoding truncated TwTPS14 codon optimised for expression in <i>Saccharomyces cerevisiae</i>
SEQ ID NO:26	Amino acid sequence of DXS of <i>Coleus forskohlii</i>
SEQ ID NO:27	Amino acid sequence of GGPPS of <i>Coleus forskohlii</i>
SEQ ID NO:28	Amino acid sequence of TPS1 of <i>Marrubium vulgare</i>

Examples

5 The invention is further illustrated by the following examples, which however, should not be construed as limiting for the invention.

Example 1

10 Full length cDNAs encoding 9 class II diTPS and 9 class I diTPS were cloned from a library of full length cDNAs. Sequences of cDNAs were determined by deep sequencing according to standard methods and putative diTPS were selected based on phylogeny essentially as described in Zerbe, Hamberger et al. 2013.

The 9 class II diTPSs catalyse formation of 6 structurally and stereochemically distinct diterpene pyrophosphate intermediates (see figure 3). The 9 class I diTPSs convert the diterpene pyrophosphate intermediates to the diterpenes. When these enzymes are expressed heterologously in *E. coli*, yeast or the *Nicotiana benthamiana*/Agrobacterium systems in combinations of specific class II and class I enzymes, it was found that even combinations of diTPS class II and class I enzymes not found in nature, would lead to production of at least 47 individual diterpenes including previously described and novel diterpenes. The individual diterpenes were detected with GC-MS and LC-MS in extracts derived from the cells overexpressing the diTPS as described below.

Transient expression in N. Benthamiana

Putative diTPS enzymes were expressed using the previously described pCAMBIA130035Su vector. pCAMBIA130035Su containing nucleic acids encoding putative diTPS and T-DNA expression plasmid containing the anti-post transcriptional gene silencing protein p19 (35S:p19)(Voinnet, Rivas et al. 2003), were transformed into the AGL-1 - GV3850 Agrobacterium strain by electroporation using a 2mm electroporation cuvette in a Gene Pulser (Bio-Rad; Capacity 25 μ F; 2.5 kV; 400 Ω). The transformed agrobacteria were subsequently transferred to 1mL YEP (yeast extract peptone) media and grown for 2-3 hours at 30°C in YEP media. 200 μ L were transferred to YEP-agar solid media containing 35 μ g/mL rifampicillin, 50 μ g/mL carbencillin and 50 μ g/mL kanamycin and grown for 2 days. Multiple colonies were transferred from the plate to 20mL YEP media in falcon tube containing 17.5 μ g/mL rifampicillin, 25 μ g/mL carbencillin and 25 μ g/mL kanamycin and grown at 30°C over night (ON) at 225 rpm. Agrobacteria were spun down and by centrifugation at 3500xg for 10 min and resuspended in 5mL H₂O. OD₆₀₀ were measured and H₂O was added to reach an OD₆₀₀=1. 3mL of agrobacteria culture containing the plasmid with nucleic acids encoding putative diTPS class II, diTPS class I and p19 gene respectively was mixed. Controls only containing either diTPS class II, diTPS class I or p19 was mixed similarly. Each mix of agrobacteria cultures were infiltrated into independent 4-6 weeks old *N. benthamiana* plants. In total 121 independent *N. benthamiana* lines were made. Plants were grown for 7 days in greenhouse before metabolite extraction.

Extraction and GC-MS analysis

3 infiltrated leafs from each *N. benthamiana* line chosen and from each of these 2 leaf disc's (\varnothing = 3cm) were carved out and added to 1 mL n-hexane with 1ppm 1-eicosene

as internal standard (IS). The 3 replicates served as experimental replicates. Extraction was done at RT for 1 hour in an orbital shaker set at 220 rpm. Plant material was spun down and extracts were transferred to new vials. Extracts were analyzed on a Shimadzu GCMS-QP2010 Ultra using an Agilent HP-5MS column (30 m x 0.250 mm i.d., 0.25 μ m film thickness). Injection volume and temperature was set at 1 μ L and 250 °C. GC program: 50 °C for 2 min, ramp at rate 4 °C min⁻¹ to 110 °C, ramp at rate 8 °C min⁻¹ to 250 °C, ramp at rate 10 °C min⁻¹ to 310 °C and hold for 5 min. Both He and H₂ were used as carrier gas and hence the retentions times were normalized with Kovat's retention index using 1ppm C₇ - C₃₀ Saturated Alkanes as reference. Electron impact (Ei) was used as ionization method in the mass spectrometer (MS) with the ion source temperature set to 230 °C and 70 eV. MS spectra's was recorded from 50m/z to 350m/z. Compound identification was done by comparison to authentic standards and comparison to reference spectra databases (Wiley Registry of Mass Spectral Data, 8th Edition, July 2006, John Wiley & Sons, ISBN: 978-0-470-04785-9). Identification was also done by C13-NMR (see below). 47 different diterpenes listed in table 1 were detected. Some of the results are also shown in figures 6 and 7. Each compound was assigned a number, and the spectrum of some of the compounds is shown in figure 6. The compound number provided in table 1 corresponds to the compound number provided figures 2 and 6. Figure 2 shows the compound names, structures and numbers. Qualitative quantification was based on the average of the experimental replicates of the total ion chromatogram (TIC) peak area normalized to the TIC area of IS.

25 Semi large scale production of miltiradiene and kovalool for NMR analysis.

For the accumulation of 0.5 - 1.5 mg of diterpene for structural analysis with NMR the diTPS class II and diTPS class I combination, which yielded the compound of interest were selected (see figure 2B). 500 mL agrobacterium cultures containing plasmids with the p19, CfDXS, CfGGPPs, diTPS class II and diTPS class I gene respectively, were grown ON from 20 mL starter cultures. All agrobacteria lines were spun down and resuspended in H₂O with to an OD₆₀₀=0.5. Whole *N. benthamiana* plants were submerged in the agrobacteria mix described above and infiltration was subsequently done by applying -70 kPa vaccum for 30 sec, similar to the method described in (Sainsbury, Saxena et al. 2012). After 7-8 days of growth leafs were harvested and "chopped". Extractions were done by 0.5L n-hexane per 100 g fresh weight leaf material. Extraction volume was reduced by rotor evaporation (Buchi, Switzerland)

set to 35°C and 220 mbar. Residual material was removed to a second vial whereas the n-hexane was reused for a repeated extraction. Extraction was repeated three times. Concentrated plant extract was applied on a Dual Layer Florisil/Na₂SO₄ 6mL PP SPE TUBE, Superleco Analytical. Elution from the column was done with a gradient eluent of n-hexane and 1-15% ethyl acetate. This was repeated 3-5 times. Fractions were analyzed with GC-MS to identify the fraction containing the diterpene of interest. Purification of miltiradiene was subsequently done on a preparative GC-MS. NMR analysis of miltiradiene was done on a Bruker 400MHz NMR instrument.

Table 2A: ¹H-NMR for the identification of miltiradiene

	(Gao, Hillwig et al. 2009)	This work
#C	δH (ppm)	δH (ppm)
7	1,896 (d), 1,931 (d)	1,993 (d), 1,929 (d)
8		
9		
10		
11	2,396 (t), 2,475 (t)	2,391 (t), 2,466 (t)
12	5,4335 (d)	5,42 (br. s)
13		
14	2,612 (2H, br. s)	2,6 (m)
15	2,159 (m)	2,156 (m)
16	0,926 (3H, d J=2.5)	0,98 (3H, d J=2.5)
17	0,999 (3H, d J=2.5)	1 (3H, d J=2.5)
18	0,8472 (3H, s)	0,84 (3H, s)
19	0,871 (3H,s)	0,87 (3H,s)
20	0,976 (3H, s)	0,97 (3H,s)

HPLC-HRMS-SPE-NMR analysis of kolavelool

The HPLC-HRMS-SPE-NMR system consisted of an Agilent 1200 chromatograph comprising quaternary pump, degasser, thermostatted column compartment, autosampler, and photodiode array detector (Santa Clara, CA), a Bruker micrOTOF-Q II mass spectrometer (Bruker Daltonik, Bremen, Germany) equipped with an electrospray ionization source and operated via a 1:99 flow splitter, a Knauer Smartline K120 pump for post-column dilution (Knauer, Berlin, Germany), a Spark Holland Prospekt2 SPE unit (Spark Holland, Emmen, The Netherlands), a Gilson 215 liquid handler equipped with a 1-mm needle for automated filling of 1.7-mm NMR tubes, and a Bruker Avance III 600 MHz NMR spectrometer (¹H operating frequency 600.13 MHz) equipped with a Bruker SampleJet sample changer and a cryogenically cooled gradient

inverse triple-resonance 1.7-mm TCI probe-head (Bruker Biospin, Rheinstetten, Germany). Mass spectra were acquired in positive ionization mode, using drying temperature of 200 °C, capillary voltage of 4100 V, nebulizer pressure of 2.0 bar, and drying gas flow of 7 L/min. A solution of sodium formate clusters was automatically
5 injected in the beginning of each run to enable internal mass calibration. Cumulative SPE trapping of kolavelool was performed after 10 consecutive separations using a chromatographic method as follows: 0 min., 90% B; 15 min., 100% B; 20 min., 100% B; 25 min., 100% B; 26 min., 90% B with 10 min. equilibration prior to injection of 5 µL pre-fractionated sample (8.5 mg/mL in hexane). The HPLC eluate was diluted with
10 Milli-Q water at a flow rate of 1.0 mL/min prior to trapping on 10 × 2 mm i.d. Resin GP (general purpose, 5-15 µm, spherical shape, polydivinyl-benzene phase) SPE cartridges from Spark Holland (Emmen, The Netherlands), and kolavelool was trapped using threshold of an extracted ion chromatogram (m/z 273.2 corresponding to $[M+H-H_2O]^+$). The SPE cartridge was dried with pressurized nitrogen gas for 60 min prior to
15 elution with chloroform-*d*. The HPLC was controlled by Bruker Hystar version 3.2 software, automated filling of NMR tubes were controlled by PrepGilsonST version 1.2 software, and automated NMR acquisition were controlled by Bruker IconNMR version 4.2 software. NMR data processing was performed using Bruker Topspin version 3.2 software.

20

NMR analyses of kolavelool

NMR spectra of kolavelool was recorded in chloroform-*d* at 300 K. 1H and ^{13}C chemical shifts were referenced to the residual solvent signal (δ 7.26 and δ 77.16, respectively). One-dimensional 1H NMR spectrum was acquired in automation
25 (temperature equilibration to 300 K, optimization of lock parameters, gradient shimming, and setting of receiver gain) with 30°-pulses, 3.66 s inter-pulse intervals, 64k data points and multiplied with an exponential function corresponding to line-broadening of 0.3 Hz prior to Fourier transform. Phase-sensitive DQF-COSY and NOESY spectra were recorded using a gradient-based pulse sequence with a 20 ppm
30 spectral width and 2k x 512 data points (processed with forward linear prediction to 1k data points). Multiplicity-edited HSQC spectrum was acquired with the following parameters: spectral width 20 ppm for 1H and 200 ppm for ^{13}C , 2k x 256 data points (processed with forward linear prediction to 1k data points), and 1.0 s relaxation delay. HMBC spectrum was optimized for $^nJ_{C,H} = 8$ Hz and acquired using the following

parameters: spectral width 20 ppm for ^1H and 240 ppm for ^{13}C , 2k x 128 data points (processed with forward linear prediction to 1k data points), and 1.0 s relaxation delay. NMR spectra of syn-isopimara-9(11), 15-diene was recorded in chloroform-*d* at 300 K on a Bruker Avance III 600 MHz NMR spectrometer (^1H operating frequency 600.13 MHz) equipped with a Bruker SampleCase sample changer and a cryogenically cooled gradient 5.0-mm DCH probe-head (Bruker Biospin, Rheinstetten, Germany) in a 3.0 mm o.d. NMR tube. ^1H and ^{13}C chemical shifts were referenced to the residual solvent signal (δ 7.26 and δ 77.16, respectively). One-dimensional ^1H and ^{13}C NMR spectrum was acquired in automation (temperature equilibration to 300 K, optimization of lock parameters, gradient shimming, and setting of receiver gain) with 30°-pulses, 3.66 s inter-pulse intervals, 64k data points and multiplied with an exponential function corresponding to line-broadening of 0.3 and 1.0 Hz, respectively prior to Fourier transform. Phase-sensitive DQF-COSY and ROESY spectra were recorded using a gradient-based pulse sequence with a 7.4 ppm spectral width and 2k x 128 and 2k x 256 data points, respectively (processed with forward linear prediction to 1k data points). Multiplicity-edited HSQC spectrum was acquired with the following parameters: spectral width 16 ppm for ^1H and 165 ppm for ^{13}C , 2k x 256 data points (processed with forward linear prediction to 1k data points), and 1.0 s relaxation delay. HMBC spectrum was optimized for $^nJ_{\text{C,H}} = 8$ Hz and acquired using the following parameters: spectral width 7.9 ppm for ^1H and 221 ppm for ^{13}C , 4k x 256 data points (processed with forward linear prediction to 1k data points), and 1.0 s relaxation delay.

Table 2B: ^1H - & ^{13}C - NMR data of (+/-)-kolavelool acquired in chloroform-*d* in HPLC-HRMS-SPE-NMR mode

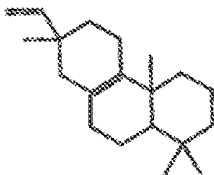
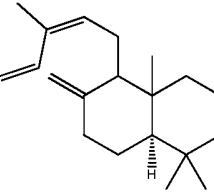
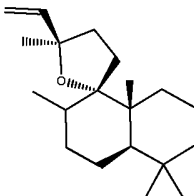
Position	(<i>Bomm, Zukerman-Schpector et al. 1999</i>)	(<i>Bomm, Zukerman-Schpector et al. 1999</i>)	This work	This work
	δ_{C}	δ_{H} (J in Hz)	δ_{C}^b	δ_{H} (J in Hz)
1	18.2		18.2	1.41 ^a 1.53 ^a
2	27.4		27	2.01 ^a
3	120.4	5.16 s	120.5	5.17, s
4	144.5		144.6	
5	38.1		37.4	
6	36.8		37.1	1.15 ^a 1.69, dt (12.0, 3.0)
7	26.8		27.6	1.40 ^a
8	36.1		36.25	1.41 ^a

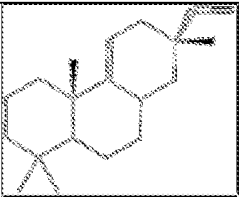
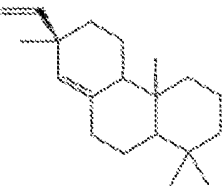
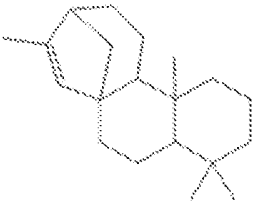
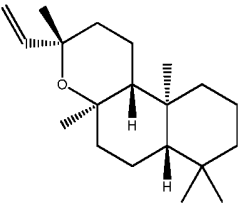
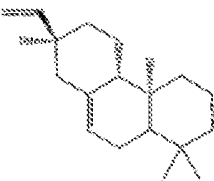
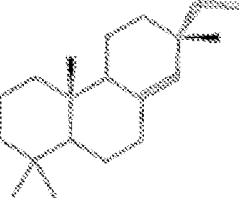
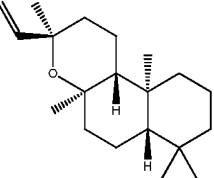
9	38.3		38	
10	46.3		46.5	1.3 ^a
11				1.38 ^a
	31.8		31.8	1.25 ^a
12	35.3		35.4	1.37 ^a
13	73.4		73.2	
14		5.84 dd (17.2, 10.8)		5.87, dd (17.4, 10.7)
	145.1		145.2	
15		5.07 dd (17.2, 1.5) 4.99 dd (10.8, 1.5)		5.04, bd (10.7) 5.18, bd(17.4)
	111.8		111.9	
16	27.7	1.24 s	27.9	1.25, s
17	15.9	0.75 d (5.9)	16	0.76, d (5.7)
18	18	1.54 d (1.5)	18	1.57, bs
19	19.2	0.95 s	20.11	0.97, s
20	18.4	0.68 s	18.5	0.71, s

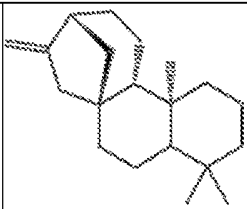
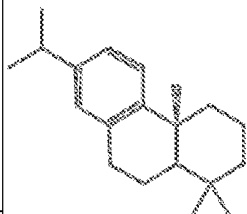
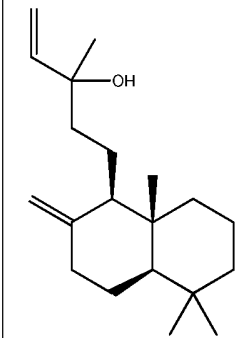
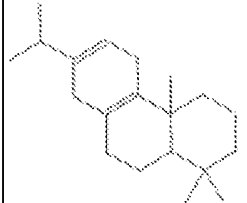
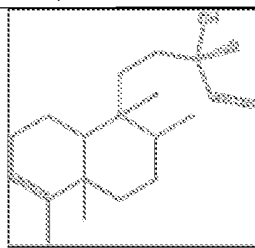
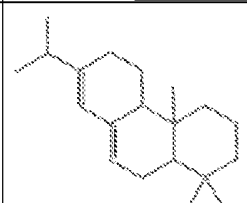
^aCoupling constants not determined due to overlap with HOD as a result of inadequate drying of cartridge in HPLC-HRMS-SPE-NMR mode; ¹H chemical shifts from HSQC experiments. ^b¹³C chemical shifts from one- and multiple-bond proton-detected 2D heteronuclear correlations.

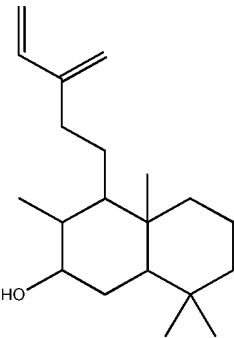
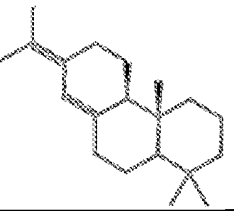
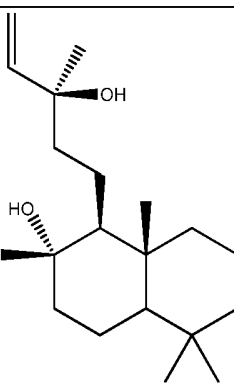
5

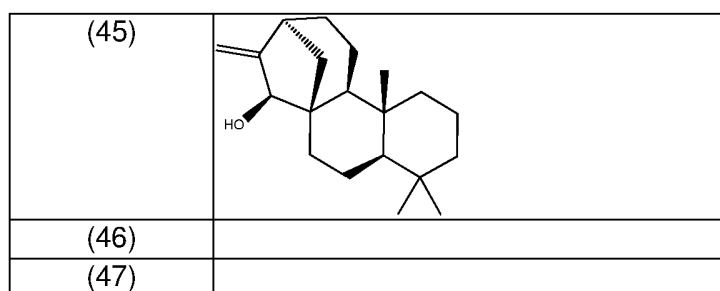
Table 1

Compound	Structure
(1)	
(2)	
(3)	
(4)	
(5)	

(6)	
(7)	
(8)	
(9)	
(10)	
(11)	
(12)	
(13)	
(14)	
(15)	
(16)	
(17)	
(19)	
(18)	
(20)	

(21)	
(22)	
(23)	
(24)	
(25)	
(26)	
(27)	
(28)	
(29)	

(30)	
(31)	
(32)	
(33)	
(34)	
(35)	
(36)	
(37)	
(38)	
(39)	
(40)	
(41)	
(42)	
(43)	
(44)	



References:

- Voinnet, O., S. Rivas, et al. (2003). "An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus." *The Plant Journal* 33(5): 949-956.
- Zerbe, P., B. Hamberger, et al. (2013). "Gene Discovery of Modular Diterpene Metabolism in Nonmodel Systems." *Plant Physiology* 162(2): 1073-1091.
- Sainsbury, F., P. Saxena, et al. (2012). Chapter Nine - Using a Virus-Derived System to Manipulate Plant Natural Product Biosynthetic Pathways. *Methods in Enzymology*. A. H. David, Academic Press. Volume 517: 185-202.

Example 2

production of syn-pimara-9,(11),15-diene (6) for NMR analysis.

15

For the structural elucidation of syn-pimara-9,(11),15-diene (6), a 0.1 L culture of a yeast strain containing OssynCPS, CfTPS3 and a GGPPs (see example 3) in a feed in time media was inoculated with a 5 mL ON culture. The culture was grown for 72 hours and harvested by adding 0.1 L of ethanol, mixing and heating to 70 °C for 20 min. After heating 0.1 L n-hexane was added, followed by horizontal shaking at 200 rpm for 1 hour. Subsequently the hexane overlay was transferred to the rotor evaporator where the volume was reduced.

Purification of syn-pimara-9,(11),15-diene (6) by solid phase extraction and preparative GC-MS.

Concentrated hexane extract from yeast was applied on a Dual Layer Florisil/Na₂SO₄ 6mL PP SPE TUBE, Superleco Analytical. Elution from the column was done with a gradient eluent of n-hexane and 1-15% ethyl acetate. This was repeated 3-5 times.

30

Fractions were analyzed with GC-MS to identify the fraction containing the diterpene of interest, these were pooled and solvent was removed by rotor evaporation and resuspended in 1 mL n-hexane. Final purification was done on an Agilent 7890B GC installed with an Agilent 5977A inert MSD, GERSTEL Preparative Fraction Collector (PFC) AT 6890/7890 and a GERSTEL CIS 4C Bundle injection port. For separation by GC a RESTEK Rtx-5 column (30m x 0.53mm ID x 1 μ m df) with H₂ as carrier gas was used. At the end of this column a split piece with a split of 1:100 to the MS and the PFC, respectively. Sufficient amount of diterpene product for NMR analysis (0.5-1 mg) was obtained by 130 injection of 5 μ L of extract. Injection port was put in solvent vent mode with 100 mL until 0.17 min. Injection temperature was held at 40 °C for 0.1 min followed by ramping at 12 °C/sec until 320, which was held for 2 min. The GC program was set to hold at 60 °C for 1 min, ramp 30 °C/min to 220 °C, ramp 2 °C/min to 250 °C and a final ramp of 30 °C/min to 220 °C, which was held for 2 min. Temperature of the transfer line from GC to PFC and the PFC itself was set to 250 °C. The PFC was set to collect the peak of syn-pimara-9,(11),15-diene (6) by their retention time identified by the MS. The method for NMR analysis for structural characterization of syn-pimara-9,(11),15-diene (6) was the same as for the analysis of kovalool (see example 1)

20

Table 3: NMR data of syn-isopimara-9(11), 15-diene^a acquired in chloroform-d

position	(<i>Oikawa, Toshima et al. 2001</i>)	This work	This work
	δ_H (J in Hz)	δ_C	δ_H (J in Hz)
1		37.8	1.36, m 1.65, m
2		19.2	1.53, m 1.65, m
3		42.5	1.16, td (13.6, 3.9) 1.40, m
4		33.8	
5		53.9	0.95, dd (12.3, 2.6)
6		22.12	1.46, m 1.66, m
7		36.4	1.01, m 1.89, m
8		31.3	2.28, m
9		149.9	

10		39.4	
11	5.29, m	112.6	5.27, ddd (6.1, 2.0, 1.5)
12		37.5	1.72, m 2.05, ddd (17.1, 2.8, 2.0)
13		34.9	
14		42.8	1.10, dd (12.6, 10.9) 1.50, m
15	5.77, dd (17.2, 11.2)	150.5	5.82, dd (17.5, 10.8)
16	4.85–4.93, m	109.3	4.87, dd (10.8, 1.4) 4.94, dd (17.5, 1.4)
17	0.95, s	22.2	0.92, s
18^b	0.84, s	33.5	0.85, s
19^b	0.84, s	22.09	0.86, s
20	0.98, s	21.1	1.04, s

^arelative stereochemistry concluded on the basis of NOE correlations between H-8 – H-20 and H-8 – H-17 as well as the absence of correlations between H-5 and H-20.

^bInterchangeable

5 Example 3

Construction of yeast strain for the production of diterpenes

Materials and methods.

10 Table 4 summarises the coding DNA sequences (CDS) used in this study. The CDS encodes the proteins indicated in Table, but have been sequence optimized for expression in yeast.

Table 4. CDSs used in this study.

CDS	Description
CfTPS1	SEQ ID NO:19 - encodes CfTPS1 (Coleus forskohlii diterpene synthase 2) truncated to remove putative plastid targeting sequence
CfTPS3	SEQ ID NO:20 – encodes CfTPS3 (Coleus forskohlii diterpene synthase 3) truncated to remove putative plastid targeting sequence
ZmAN2	SEQ ID NO:21 – encodes ZmAN2 (Zea Maiz diterpene synthase class II) truncated to remove putative plastid targeting sequence
OssynCPS	OssynCPS (Oryza sativa diterpene synthase class II) truncated to remove putative plastid targeting sequence
TwTPS21	SEQ ID NO:23 – encodes TwTPS21 (Tripterygium wilfordii diterpene synthase class II) truncated to remove putative plastid targeting sequence

SsSCS	SEQ ID NO:24 – encodes SsSCS (Salvia Sclarea diterpene synthase class I) truncated to remove putative plastid targeting sequence
TwTPS14	SEQ ID NO:25 – encodes TwTPS14 (Tripterygium Wilfordii diterpene synthase class II) truncated to remove putative plastid targeting sequence
GGPPs	Geranylgeranyl diphosphate synthase

Table 5. List of plasmids used in the study.

pCYPCC-1	pROP196 XI-5 assembler 1	Rv #205 GGPPs7<-pTPI1 #219
pCYPCC-2	pROP196 XI-5 assembler 1	Rv #206 GGPPs10<-pTPI1 #219
pCYPCC-3	pROP196 XI-5 assembler 1	Rv #205 GGPPs7<-pPGK1 1c
pCYPCC-4	pROP196 XI-5 assembler 1	Rv #206 GGPPs10<-pPGK1 1c
pCYPCC-7	pROP197 XI-5 assembler 3	#-3 CfTPS3 <-#161pTDH3
pCYPCC-9	pVAN858 assembler 2	2c pTEF1->#-5 CfTPS1
pCYPCC-10	pVAN858 assembler 2	2c pTEF1->#-6 OsCPssyn
pCYPCC-18	pROP197 XI-5 assembler 3	#-8 SsSCS <-#161pTDH3
pCYPCC-21	pROP197 XI-5 assembler 3	Res# 236 CfTPS3 co<-#161pTDH3
pCYPCC-42	pVAN858 assembler 2	Res160 pTEF-2 ->CfTPS1, co
pCYPCC-44	pVAN858 assembler 2	Res160 pTEF-2 ->OsCPssyn
pCYPCC-51	pROP197 XI-5 assembler 3	SsSCS, co<-#161pTDH3

All enzymes cloned in plasmids pCYPCC7-51 were truncated to remove putative plastid targeting sequence (see sequence listing).

- Abbreviation: co=codon optimized. Codon optimization for *Saccharomyces cerevisiae* was performed using the Genart service from LifeTechnologies.
- DNA fragments containing the enzymes of interest were USER cloned into pre-digested plasmid backbones. All plasmids constructed and used in this study are summarized in table 5. DNA fragments of interest were liberated from plasmids by NotI enzyme-digestion as linear DNA fragments suitable for yeast transformation. The plasmids are designed to accommodate integration of up to three NotI-digested fragments at the same site in the genome.

Table 6. Strains used and generated in this study

Strain	CDS	Compound produced	Analysis
T2	TwTPS14+SsSCS+GGPPs	Kovalool (26)	GC-MS
T5	ZmAN2+SsSCS+GGPPs	ent-manool (23b)	GC-MS /LC-MS
T8	TwTPS21+EpTPS1+GGPPs	13S-manoyl oxide (20)	GC-MS
EFSC4725	CfTPS1+SsSCS+GGPPs	(+)-manool	GC-MS/ LC-MS
EFSC4727	OssynCPS+SsSCS+GGPPs	syn-manool (11)	LC-MS
EFSC4690	OssynCPS+CfTPS3+GGPPs	syn-pimara-9,(11),15-diene (6), syn-isopimara-7,15-diene (19)	GC-MS
EFSC4691	CfTPS1+CfTPS3+GGPPs	Miltiradiene (25)	GC-MS
EFSC4494	CfTPS2+CfTPS3+GGPPs	13R-manoyl oxide	GC-MS

- All strains were grown in 96 deep well plates as follows. Single colonies were inoculated in 500 µl SC-Ura in 2.2 ml 96 deep well plates and grown o/n @ 30°C, 400 RPM. The following day 50 µl of the o/n culture was used as inoculum in 500 µl DELFT media with 10% sun flower oil and grown for additional 72 hours @ 30°C, 400 RPM.
- Table 6 summarizes the compounds produced by the various strains. The table also indicates whether the compound was identified LC-MS and/or GC-MS. LC-MS analysis and/or GC-MS analysis were performed as described below. The numbers indicated in brackets refer to the compounds numbers shown in figure 2.

Extraction and LC-MS analysis

Metabolites were extracted from the whole broth by adding 500 µl 96 % Ethanol, mix and incubate @ 78°C for 10 min. For LC-MS analysis cell debris was removed by centrifugation for 2 min at 15000 xg. Supernatant was used for LC-MS analysis. LC-MS was carried out using an Agilent 1100 Series LC (Agilent Technologies, Germany) coupled to a Bruker HCT-Ultra ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany). A Zorbax SB-C18 column (Agilent; 1.8 µm, 2.1 x 50 mm) maintained at 35°C was used for separation. The mobile phases were: A, water with 0.1% (v/v) HCOOH and 50mM NaCl; B, acetonitrile with 0.1% (v/v) HCOOH. The gradient program was: 0 to 1 min, isocratic 50% B; 1 to 10 min, linear gradient 50 to 95% B; 10 to 11.4 min, isocratic 98% B; 11.4 to 17 min, isocratic 50% B. The flow rate was 0.2 mL min⁻¹. The mass spectrometer was run in alternating positive/negative mode and the range m/z 100-800 was acquired.

Extraction GC-MS analysis

Metabolites were extracted from the whole broth by adding 500 µl 96 % Ethanol, mix and incubate @ 78°C for 10 min. Solvent and liquids were removed by freeze drying. 500 µL of hexane including 1mg/L 1-eicosene as internal standard (ISTD), was used for extraction at room temperature for ½ an hour. Particles in the extraction media was removed by centrifugation for 2 min at 15000 xg. After extraction, the solvent was transferred into new 1.5-mL glass vials and stored at -20 °C until GC-MS analysis. One microliter of hexane extract was injected into a Shimadzu GC-MS-QP2010 Ultra. Separation was carried out using an Agilent HP-5MS column (20m 0.180mm i.d., 0.18µm film thickness) with purge flow of 4 mL min⁻¹ for 1 min, using H₂ as carrier gas. The GC temperature program was 60°C for 1 min, ramp at rate 30°C min⁻¹ to 180°C, ramp at rate 10°C min⁻¹ to 250°C, ramp at rate 30°C min⁻¹ to 320°C, and hold for 3 min. Injection temperature was set at 250°C in splitless mode. Column flow and pressure was set to 5. mL min⁻¹ and 66.7 kPa yielding a linear velocity of 66.5 cm s⁻¹. Ion source and transfer line for mass spectrometer (MS) was set to 300°C and 280°C respectively. MS was set in scan mode from m/z 50 to m/z 350 with a scan width of 0.5s. Solvent cutoff was 4 min.

Claims

1. A method of producing a terpene, said method comprising the steps of:

5 a) providing a host organism comprising

I. A heterologous nucleic acid encoding a diTPS of class II,

II. A heterologous nucleic acid encoding a diTPS of class I,

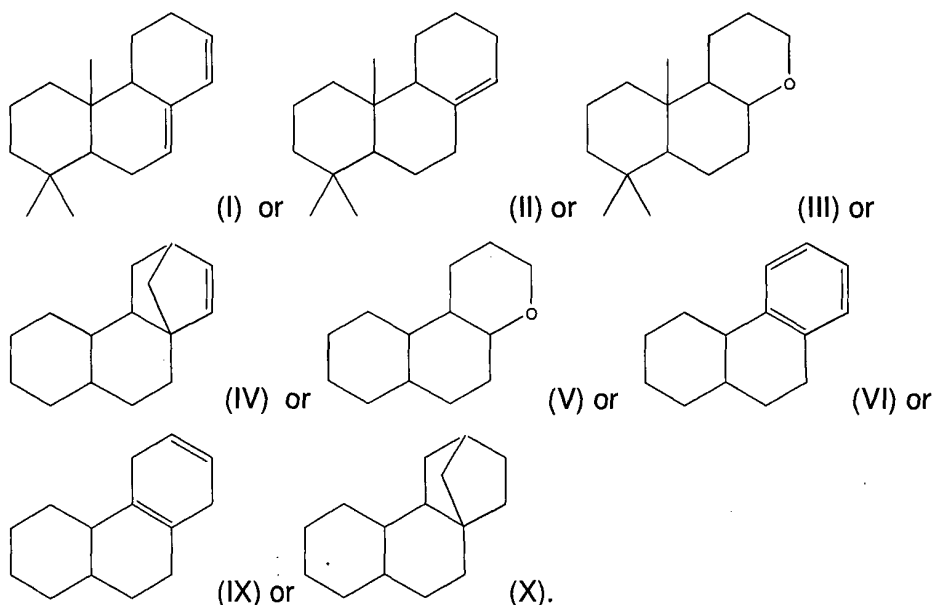
with the proviso that said diTPS of class II and said diTPS of class I is not from the same species; and with the proviso that when said diTPS of class II is SsLPPS then said diTPS of class I is not CfTPS3, CfTPS4 or EpTPS8 and when said diTPS of class I is EpTPS8, then the diTPS of class II is not CfTPS2 or SsLPPS;

b) Incubating said host organism in the presence of geranylgeranyl pyrophosphate (GGPP) under conditions allowing growth of said host organism;

15 c) Optionally isolating diterpene from the host organism.

2. The method according to claim 1, wherein diterpene is a C₂₀-molecule containing a decalin core and up to 3 oxygen molecules.

20 3. The method according to any one of claims 1 to 2, wherein the diterpene is a C₂₀-molecule containing a core structure of one of following formulas I, II, III, IV, V, VI, IX or X:



25

4. The method according to claim 3, wherein diterpene is a C₂₀-molecule containing a cores structure of formulas I, II, III, IV, V, VI, IX or X substituted at one or more positions by one or more groups selected from the group consisting of:
- 5 a) alkyl, such as C₁₋₆-alkyl, for example C₁₋₃, wherein said alkyl may be linear or branched, for example alkyl may be isopropyl or methyl
- b) alkenyl, such as C₁₋₆ alkenyl, such as C₂₋₄-alkenyl, such as C₂₋₃-alkenyl; and
- c) hydroxyl.
- 10 5. The method according to any one of claims 1 and 2, wherein the diterpene is a C₂₀-molecule containing a decalin substituted at the 10 position with C₅-alkenyl chain, which optionally may be substituted with a hydroxyl and/or a methyl group and/or =C.
- 15 6. The method according to any one of claims 1 and 2, wherein the diterpene is a C₂₀-molecule consisting of 20 carbon atoms, up to three oxygen atoms and hydrogen atoms, and which contains a core structure of any of formulas I, II, III, IV, VI, X, XXII, XXIII, XXIV, XXV, XXVI, XXVII, XXVIII, XXIX, XXX, XXXI, XXXII, XXXIII, XXXIV, XXXV, XXXVI, XXXVII, XXXVIII, XXXIX, XL and/or XLI.
- 20 7. The method according to claim 1, wherein the diterpene is the product of any of the reactions VII to XIX described herein above.
8. The method according to claim 1, wherein the diterpene is any of the compounds 1 to 47 of Table 1.
- 25 9. A host organism comprising
- I. A heterologous nucleic acid encoding a diTPS of class II;
- II. A heterologous nucleic acid encoding a diTPS of class I,
- with the proviso that said diTPS of class II and said diTPS of class I is not
- 30 from the same species.
10. The method according to any one of claims 1 to 8, or the host organism according to claim 9, wherein the diTPS of class II is a polypeptide sharing at least 30%, such as at least 35% sequence identity to the sequence of SsLPPS (SEQ ID NO:6) or to the sequence of AtCPS of figure 5.

11. The method or the host organism according to claim 10, wherein the said diTPS of class II contains the following motif of four amino acids:

5 D/E-X-D-D, wherein X may be any amino acid .

12. The method according to any one of claims 1 to 8 and 10 to 11 or the host organism according to any one of claims 9 to 11, wherein the diTPS of class II is selected from the group consisting of syn-CPP type diTPS, ent-CPP type diTPS, (+)-
10 CPP type diTPS, LPP type diTPS and LPP like type diTPS.

13. The method according to any one of claims 1 to 8 and 10 to 12, or the host organism according to any one of claims 9 to 12, wherein the diTPS of class I is a polypeptide sharing at least 30%, such as at least 35% sequence identity to the
15 sequence of ScSCS (SEQ ID NO:11) or to the sequence of AtEKS of figure 4.

14. The method or the host organism according to claim 13, wherein the said diTPS of class I contains the following motif of five amino acids:

20 D-D-X-X-D/E, wherein X indicates any amino acids.

15. The method according to any one of claims 1 to 8 and 10 to 14 or the host organism according to any one of claims 9 to 14, wherein the diTPS of class I is selected from the group consisting of EpTPS8 like diTPS, EpTPS23 like diTPS, SsSCS
25 like diTPS, CfTPS3 like diTPS, CfTPS4 like diTPS, MvTPS5 like diTPS, TwTPS2 like diTPS, EpTPS1 like diTPS and CfTPS14 like diTPS.

16. A polypeptide, which is EpTPS7 of SEQ ID NO:2 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least
30 90%, such as at least 95% sequence identity therewith.

17. A polypeptide, which is TwTPS7 of SEQ ID NO:4 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least
35 90%, such as at least 95% sequence identity therewith.

18. A polypeptide, which is CfTPS1 of SEQ ID NO:5 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.
- 5 19. A polypeptide, which is TwTPS21 of SEQ ID NO:7 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.
- 10 20. A polypeptide, which is TwTPS14/28 of SEQ ID NO:8 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.
- 15 21. A polypeptide, which is EpTPS23 of SEQ ID NO:10 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.
- 20 22. A polypeptide, which is TwTPS2 of SEQ ID NO:14 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.
- 25 23. A polypeptide, which is EpTPS1 of SEQ ID NO:15 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.
- 30 24. A polypeptide, which is CfTPS14 of SEQ ID NO:16 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.
- 35 25. The method according to any one of claims 1 to 8 and 10 to 15, or the host organism according to any one of claims 9 to 15, wherein the diTPS of class II is the polypeptide according to any one of claims 16 to 20.
26. The method according to any one of claims 1 to 8 and 10 to 15, or the host organism according to any one of claims 9 to 15, wherein the diTPS of class II is an enzyme capable of catalysing any of the reactions I to V.

27. The method according to any one of claims 1 to 8 and 10 to 15, or the host organism according to any one of claims 9 to 15, wherein the diTPS of class II is syn-CPP of SEQ ID NO:1 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.

28. The method according to any one of claims 1 to 8 and 10 to 15, or the host organism according to any one of claims 9 to 15, wherein the diTPS of class II is ZmAN2 of SEQ ID NO:3 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith.

29. The method according to any one of claims 1 to 8 and 10 to 15, or the host organism according to any one of claims 9 to 15, wherein the diTPS of class II is SsLPPS of SEQ ID NO:6 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith, with the proviso that the diTPS of class I is not ScSCS, CfTPS3, CfTPS4 or EpTPS8.

30. The method according to any one of claims 1 to 8 and 10 to 15, or the host organism according to any one of claims 9 to 15, wherein the diTPS of class II is CfTPS2 of SEQ ID NO:17 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith, with the proviso that the diTPS of class I is not CfTPS3, CfTPS4 or EpTPS8.

31. The method according to any one of claims 1 to 8 and 10 to 15, or the host organism according to any one of claims 9 to 15, wherein the diTPS of class II is an enzyme capable of catalysing at least one of the reactions XXXIII, XXXIV, XXXV, XXXVI.

32. The method according to any one of claims 1 to 8 and 10 to 15, or the host organism according to any one of claims 9 to 15, wherein the diTPS of class II is MvTPS1 of SEQ ID NO:28 or a functional homologue thereof sharing at least 70%,

such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith, with the proviso that the diTPS of class I is not MvTPS5.

5 33. The method according to any one of claims 1 to 8 and 10 to 15 and 25 to 32, or the host organism according to any one of claims 8 to 15 and 25 to 32, wherein the diTPS of class I is the polypeptide according to any one of claims 21 to 24.

10 34. The method according to any one of claims 1 to 8 and 10 to 15 and 25 to 32, or the host organism according to any one of claims 9 to 15 and 25 to 32, wherein the diTPS of class I is an enzyme capable of catalysing any of the reactions VII to XIX.

15 35. The method according to any one of claims 1 to 8 and 10 to 15 and 25 to 32, or the host organism according to any one of claims 9 to 15 and 25 to 32, wherein the diTPS of class I is an enzyme capable of catalysing at least one of the reactions X, XXII, XXIV, XXX, XXXI and XXXII.

20 36. The method according to any one of claims 1 to 8 and 10 to 15 and 25 to 32, or the host organism according to any one of claims 9 to 15 and 26 to 33, wherein the diTPS of class I is SsSCS of SEQ ID NO:11 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith, with the proviso that the diTPS of class II is not SsLPPS.

25 37. The method according to any one of claims 1 to 8 and 10 to 15 and 25 to 32, or the host organism according to any one of claims 9 to 15 and 25 to 32, wherein the diTPS of class I is CfTPS12 of SEQ ID NO:12 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith, with the proviso that the diTPS of class II is not CfTPS2.

30 38. The method according to any one of claims 1 to 8 and 10 to 15 and 25 to 32, or the host organism according to any one of claims 9 to 15 and 25 to 32, wherein the diTPS of class I is MvTPS5 of SEQ ID NO: 18 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as

at least 95% sequence identity therewith, with the proviso that the diTPS of class II is not MvTPS1.

5 39. The method according to any one of claims 1 to 8 and 10 to 15 and 25 to 32, or the host organism according to any one of claims 9 to 15 and 25 to 32, wherein the diTPS of class I is CfTPS3 of SEQ ID NO:12 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith, with the proviso that the diTPS of class II is not CfTPS2 or SsLPPS.

10

40. The method according to any one of claims 1 to 8 and 10 to 15 and 25 to 32, or the host organism according to any one of claims 9 to 15 and 25 to 32, wherein the diTPS of class I is EpTPS8 of SEQ ID NO:9 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith, with the proviso that the diTPS of class II is not CfTPS2 or SsLPPS.

15

41. The method according to any one of claims 1 to 8 and 10 to 15 and 25 to 32, or the host organism according to any one of claims 9 to 15 and 25 to 32, wherein the diTPS of class I is CfTPS4 of SEQ ID NO:13 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95% sequence identity therewith, with the proviso that the diTPS of class II is not CfTPS2 or SsLPPS.

20

25 42. The method or the host organism according to any one of the preceding claims, wherein the host organism further comprises one or more heterologous nucleic acids encoding enzymes involved in the biosynthesis of GGPP.

43. The method or the host organism according to any one of the preceding claims, wherein said enzymes is selected from the group consisting of CfDXS of SEQ ID NO:26, CfGGPPS of SEQ ID NO:27 and functional homologues of any of the aforementioned sharing at least 70% sequence identity therewith.

30

44. The method or the host organism according to any one of the preceding claims, wherein the host organism is a microorganism.

35

45. The method or the host organism according to claim 44, wherein the microorganism is yeast.

5 46. The method according to any one of claims 1 to 44, wherein the host organism is a plant.

47. A method of producing a diterpene, said method comprising the steps of

10 a) providing a host organism according to any one of claims 9 to 15 and 25 to 46;

b) preparing an extract of said host organism;

c) providing GGPP

d) incubating said extract with GGPP

thereby producing a diterpene.

15

48. A method for producing kolavelool comprising the steps of:

a) providing a host organism comprising

20

I. A heterologous nucleic acid encoding a diTPS of class II which is an CLPP like type diTPS;

II. A heterologous nucleic acid encoding a diTPS of class I;

b) Incubating said host organism in the presence of geranylgeranyl pyrophosphate (GGPP) under conditions allowing growth of said host organism;

c) Optionally isolating kolavelool from the host organism.

25

49. The method according to claim 48, wherein the CLPP type diTPS is capable of catalysing reaction XXXV.

30

50. The method according to claim 48, wherein the CLPP type diTPS is TwTPS14/28 of SEQ ID NO:8 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99% sequence identity therewith.

35

51. The method according to any one of claims 48 to 50, wherein the diTPS of class I is capable of catalysing reaction XXXVII.

52. The method according to any one of claims 48 to 51, wherein the diTPS of class I is SsSCS of SEQ ID NO:11 or a functional homologue thereof sharing at least 70%, such as at least 80%, such as at least 85%, such as at least 90%, such as at least 95%,
5 such as at least 98%, such as at least 99% sequence identity therewith.

10

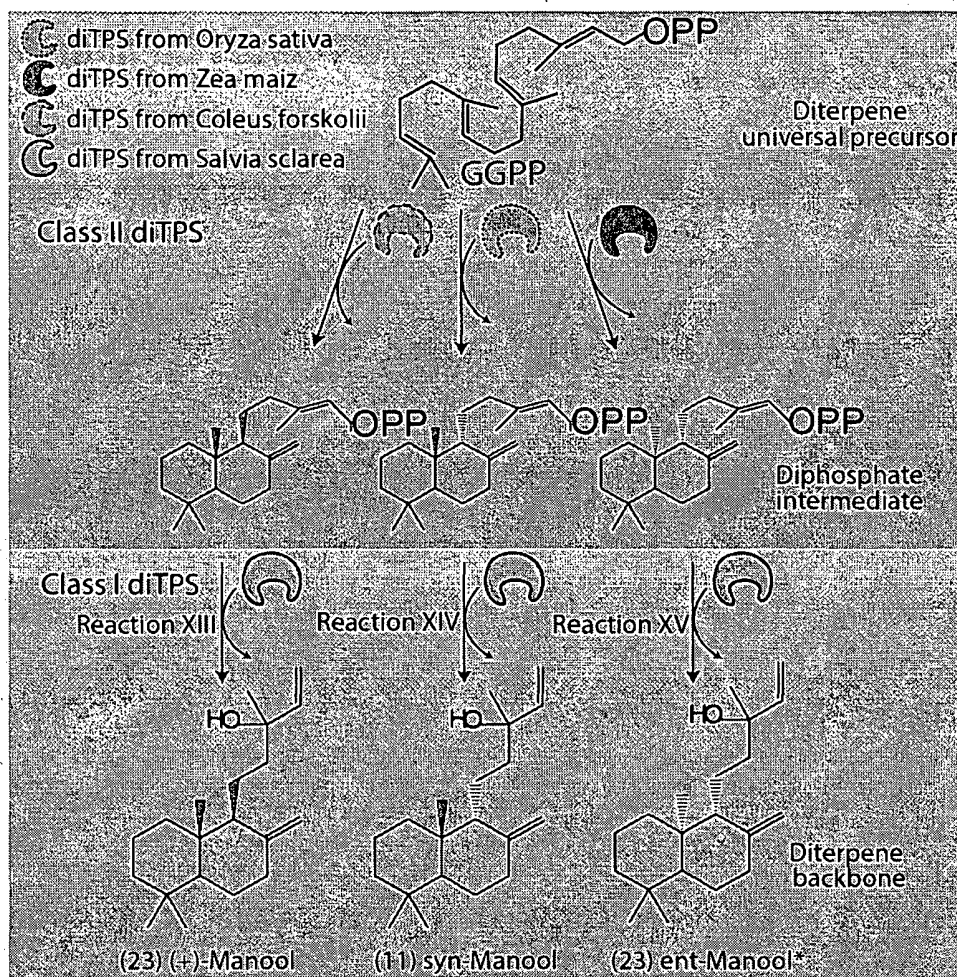


Fig. 1

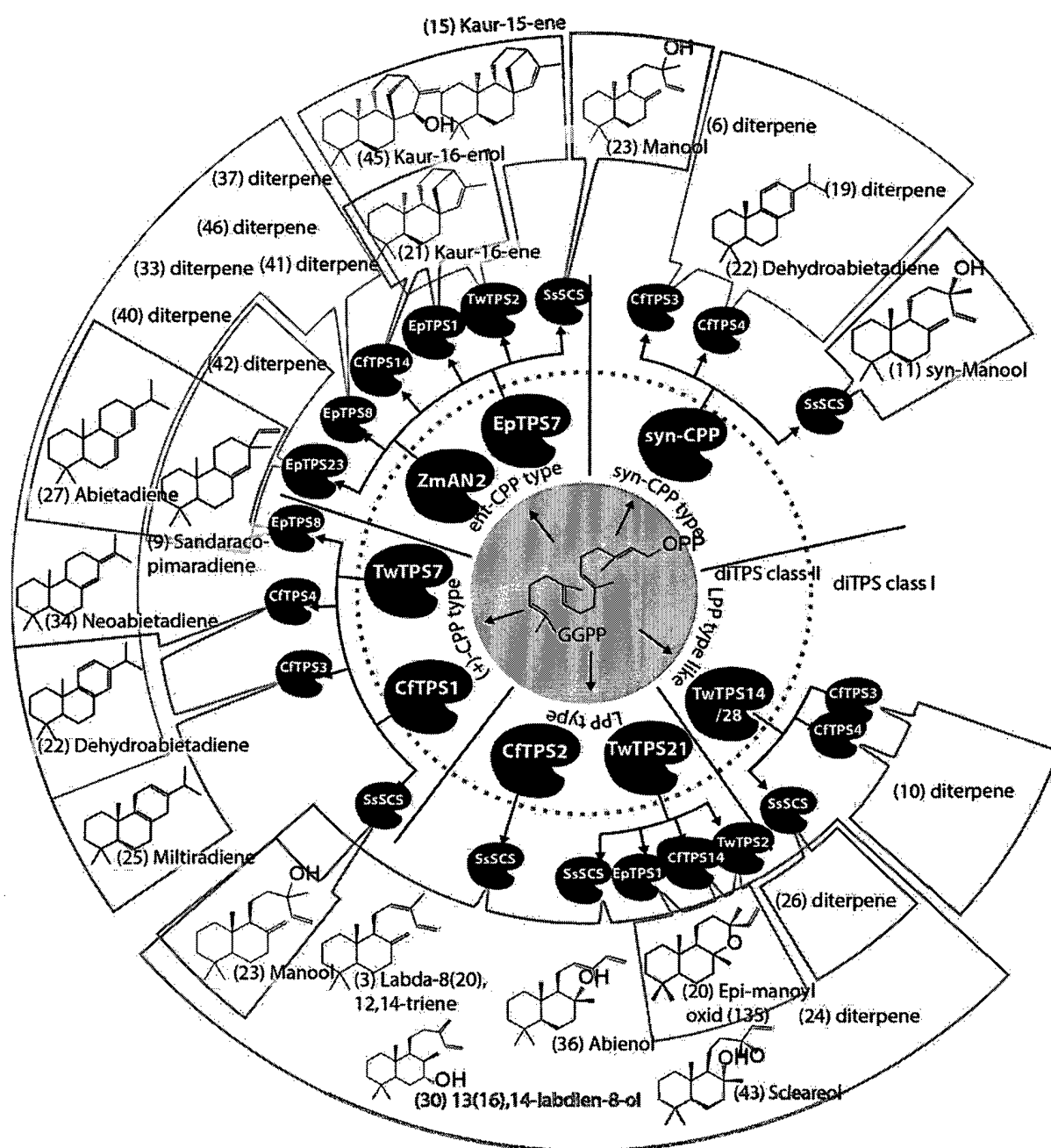


Fig. 2A

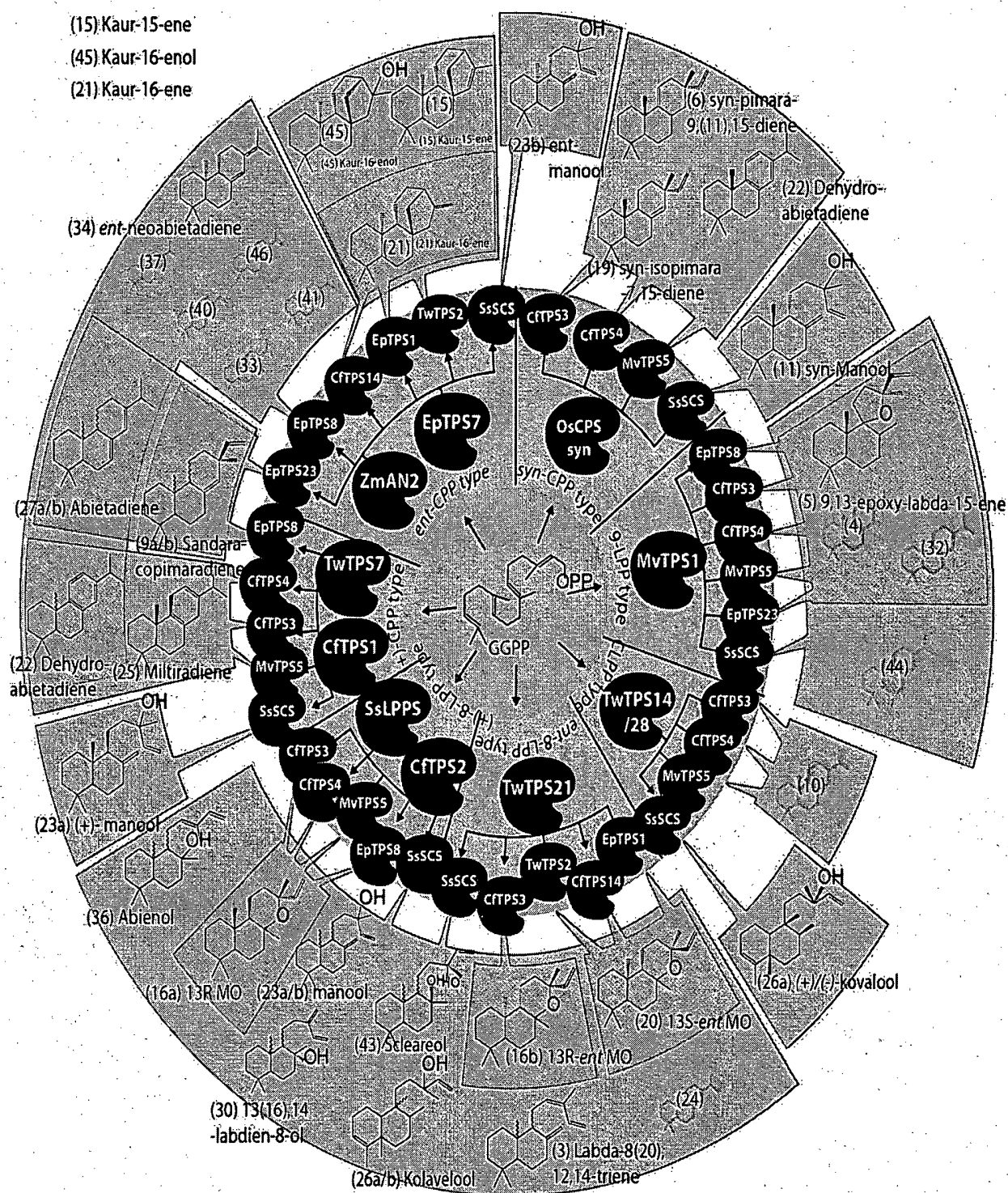
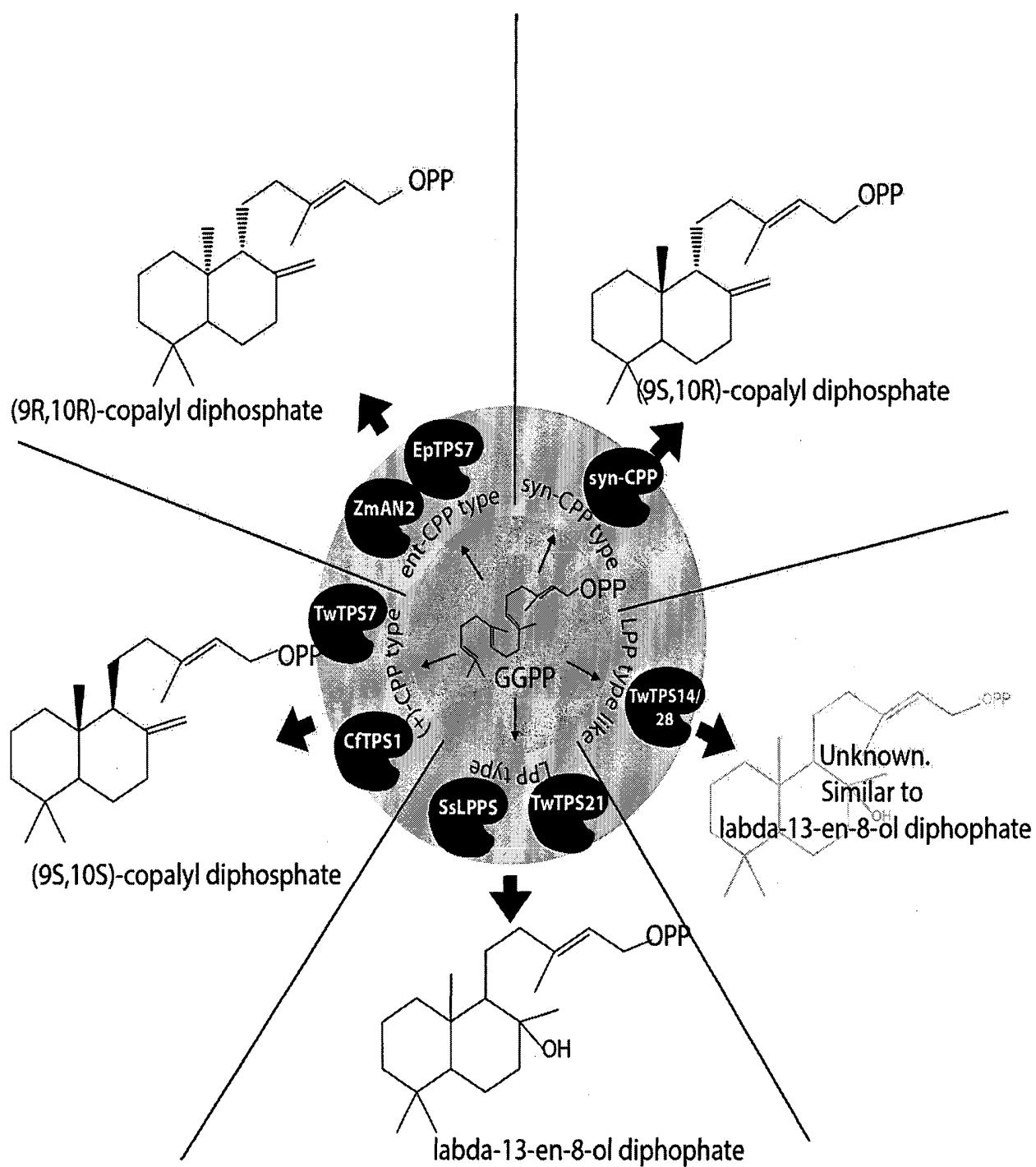
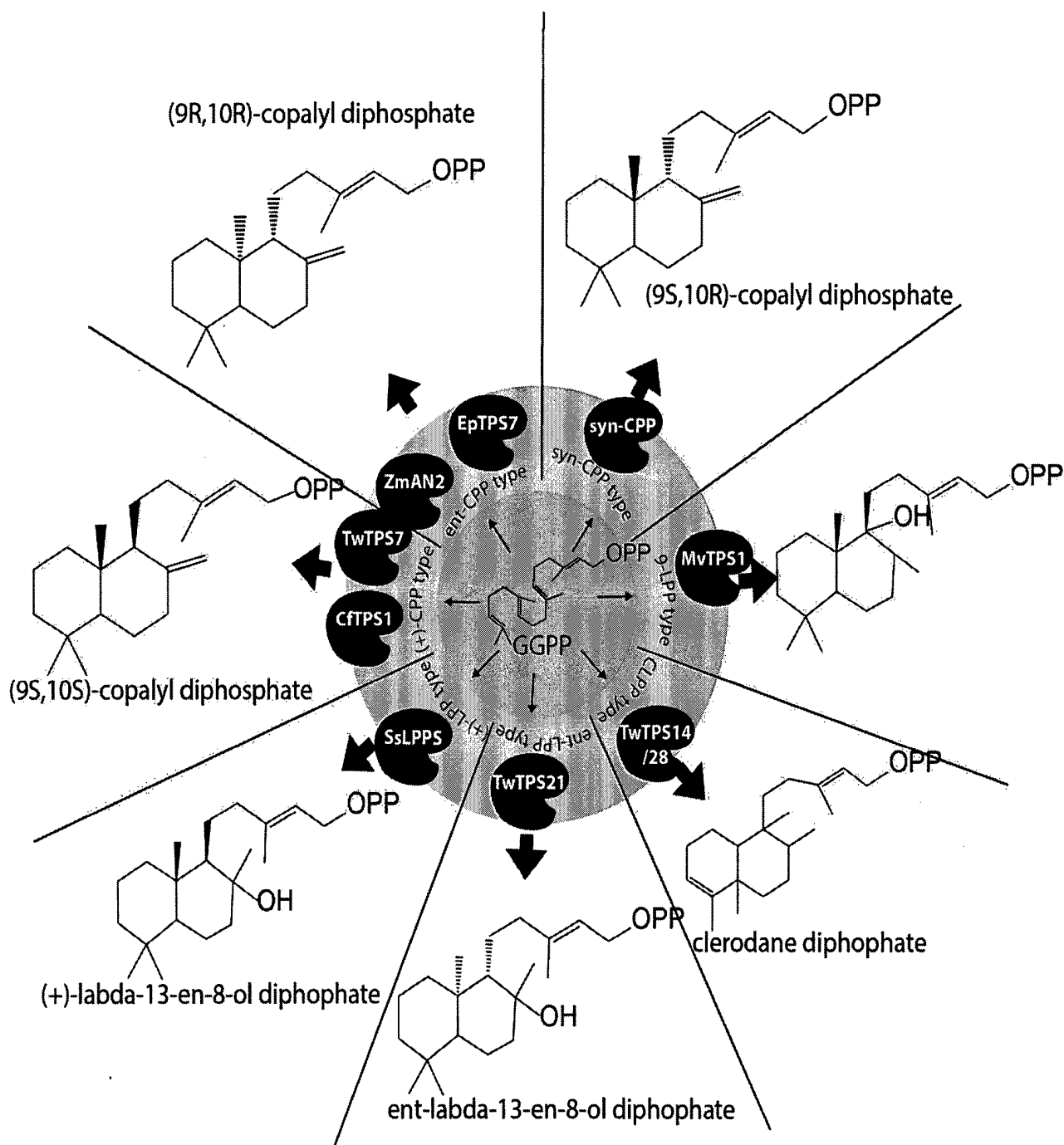


Fig. 2B

**Fig. 3A**

**Fig. 3B**

	10	20	30	40	50	60	70
CfTPS3	-----	-----	-----	-----	-----	-----	-----
CfTPS4	-----	-----	-----	-----	-----	-----	-----
EpTPS1	-----MSATP-NSFFTSP-----	-----	-----ISAKLGHPKSQSVAESNTRIQQ-LDGTREKIKKMFDRVELS				
TwTPS2	-----	-----	-----	-----	-----	-----MFDKTQLS	
CfTPS14	-----MSLPLSTCVLFVPKGSQFWSSRFSYASASLEVGFQRATSAQIAPLSKSFEETKGRIAKLFHDELS						
EpTPS8	MQVSLSLTTGSEPCITRIHAPS-----						
EpTPS23	-----MLASSTSSRFTKEWEP SNKTFSGSVRAQLSQRVKNIVVTPDQVKESESSGTSLLKEMLKRVEMP						
SsSCS	-----	-----	-----	-----	-----	-----	-----
AtEKS	-----MSINLRSSGCGSSP-----	-----	-----ISATLERGLDSEVQTRANVNS--FEQTKEKIRKMLEKVELS				
	80	90	100	110	120	130	140
CfTPS3	-----	-----	-----	-----	-----	-----MSSLAGNERVIPES	
CfTPS4	-----	-----	-----	-----	-----	-----MSLTINRVIAPF	
EpTPS1	VSPYDTAWVAMVPS-PNSLEAPVFEECSKVIQNDQNDGSGWGVY-HRDEPLIVKDSISSTLACVILALKRWG						
TwTPS2	VSAYDTAWVAMVSS-PNSRQAPWFEPCVNWLLDNDQSDGSGWGLP-PHPSPVVKDALSSSTLACVILALKRWG						
CfTPS14	ISTYDTAWVAMVPS-PTSSEEPCEPACINWLLDNDQSDGSGWARP-HHHPMLKKDVLSSSTLACVILALKRWG						
EpTPS8	TSSYDTAWVAMVPCSSNS-SKPLFPESIKWIMENONPEGNWAVDHAHPLIKDLSSTLACVILALKRWG						
EpTPS23	ISSYDTAWVAMVPSMEHSRNKPLFPNSLKWIMENONQPDGSGWCFDDSNHPWLIKDLSSTLASVILALKRWG						
SsSCS	-----	-----	-----	-----	-----	-----MSLAFNVGVTPES	
AtEKS	VSAYDTSWVAMVPS-PSSQNAFLFPQCVKWLDDNDQHDGSGWGLDNHDEHQSILKKDVLSSSTLASVILALKRWG						
	150	160	170	180	190	200	210
CfTPS3	GNRVOTRTGILPVHQTP--MITSKSSAAVKCSLITP						
CfTPS4	GHGVQSRQGLFAVMEFPRNKNTFKSSFAVKCSLSTP						
EpTPS1	IGEQVNVKGLFIEELNSASLNDLKQYKPVGDDITFPRMLEHAKDFGINLPIIDPKYVEAVIEFSDIDILKSG						
TwTPS2	IGEQVNVKGLFIEELNSASLNDLKQYKPVGDDITFPRMLEHAKDFGINLPIIDPKYVEAVIEFSDIDILKSG						
CfTPS14	VGEQVNVKGLFIEELNSASLNDLKQYKPVGDDITFPRMLEHAKDFGINLPIIDPKYVEAVIEFSDIDILKSG						
EpTPS8	LAPQLVHSGIDEIGSNLWAAAMFQRSEPTGPDVIEPGLMHQAIDLGINLPFNNSSTIENMITNPLIDIQS-						
EpTPS23	VEQQLIDKGLFIEELNSMWAATDMHQYSEIGENIIEPSPMVEHANKLGISLISIDHSLFQSMIRNRDMETKS-						
SsSCS	GQRVGSRRKEKFPVQGFV-VTPPNRSRLIVNCSLIT						
AtEKS	IGERQVNVKGLFIEELNSALVTDETIOKPTGDDITFPGMLKYARDINITIPIGSEVVDDMIRKRDIIDILKCD						
	220	230	240	250	260	270	280
CfTPS3	-----	-----	-----	-----	-----	-----	-----
CfTPS4	-----	-----	-----	-----	-----	-----	-----
EpTPS1	CDSTTEGRKAYIAYISEGIGN-IQDWNVMKYQRNGSLFSPSATAAASIHHDASCLRYTCALKKEG						
TwTPS2	---NKLEGREAYIAYVSEGIGK-IQDWNVMKYQRNGSLFSPSATAAASIHHDASCLRYTCALKKEG						
CfTPS14	YQSSSTEREVYRAYIAEGIGK-IQDWNVMKYQRNGSLFSPSATAAASIHHDASCLRYTCALKKEG						
EpTPS8	---FEAGKTSHTAYTAEGIGSRIDKEQLQYQTSNGSLFSPSATAAASIHHDASCLRYTCALKKEG						
EpTPS23	---LNG---RNMAYVAEGLNG-SNNWKEVMKYQRNGSLFSPSATAAASIHHDASCLRYTCALKKEG						
SsSCS	-----	-----	-----	-----	-----	-----	-----
AtEKS	SEKFSKGREAYIAYVLEGTRN-IQDWNVMKYQRNGSLFSPSATAAASIHHDASCLRYTCALKKEG						
	290	300	310	320	330	340	350
CfTPS3	-----	-----	-----	-----	-----	-----	-----
CfTPS4	-----	-----	-----	-----	-----	-----	-----
EpTPS1	N-AVPTIYPFNLYVRLSMVDAIESLGIAHFEQBEIKVILDETYRYWLOGNEEIFODCTTCAMAFRLLRVK						
TwTPS2	N-AVPTIYPFNLYVRLSMVDAIESLGIAHFEQBEIKVILDETYRYWLOGNEEIFODCTTCAMAFRLLRVK						
CfTPS14	D-AVPAVEPLDIYSLCTVDNLERVGISRHELTETIOSVLDCTYRSWLOGDEQIEFDASTCAAFRLLRMN						
EpTPS8	NGAVPTIYPLDARTISLIDSLEKFGISHEIQEMTIIILDQIYSEWKEGNEEIFKDPGCCATAFRLLRKH						
EpTPS23	H-AVPTIYFEDIYARLCIDELEKLGVDRFVEIEKMLILDYIYRCWLEGSEIILEDPTCCAMAFRLLRMN						
SsSCS	-----	-----	-----	-----	-----	-----	-----
AtEKS	A-AVPSVYFEDQYARLSIIVTLESIGIDRDEKTEIKSLIDEITYRYWLRGDEEICIDLATCAAFRLLLAH						

Fig 4

	360	370	380	390	400	410	420
CfTPS3	GYEVASDELAPYADQE	-----RINLQ	TIDVPTVVELYRAAQ	ERITEED	STLEKLYVWT	SAFLKQOILLTD	
CfTPS4	GYDVSSEELAPYADQE	-----GMNLQ	TIDILAAVIELYRAAQ	ERVAEED	STLEKLYVWT	STFLKQOILLAG	
EpTPS1	GYNVSSSEKLNQFTEDH	-FSN--	SLGGYLEDMRPVLELYKASQ	-LIFPDEL	FLEKQF	SWTSOCLKOKISSG	
TwTPS2	GYEVSSDPLTQCAHH	-FSR--	SFGGHLKDFSTALELKASQ	FVIFPEES	GLEKQMSWT	NOFLKQFNSG	
CfTPS14	GYNVSSDPIITKLIQEG	-----SF	SRNTMDINTTILELYRASEL	ILYPDER	DLEEHNLR	LKTILOELSGG	
EpTPS8	GYDVSSDSLAEFEKKEIF	YHSSA	ASAHEIDTKSILELFRASOMK	ILONEP	IILDRYDWT	SIFLRDQLVKG	
EpTPS23	GYVVSVDVLQGFEEEEKLFH	-----VKDTK	SVLELLKASQ	LKVSEKE	GILDRIYSW	ATSYLKHOLFNA	
SsSCS	GYEVSSSEELAPYGNQE	-----AVSQO	TNDLPMIELYRAANER	IYEEERS	SLEKILAWT	TIFLNKQVQDN	
AtEKS	GYDVSYPDLKPF	AEESGFS	D--TLEGYVKN	TFVLELEKAAQ	--SYPHES	ALKKQCCWTKOYLEMELSS--	
	430	440	450	460	470	480	490
CfTPS3	-----AIPDKKLHKOVE	YLYKNY--	HGILDRMGVRRNIDL	-YDISHYKSL	KAHRFYNLS	NE--DILAFARO	
CfTPS4	-----AIPDQKLHKOVE	YLYKNY--	HGILDRMGVRKGLDL	-YDAGYKALK	KAADRLVDL	CNE--DILAFARO	
EpTPS1	-LRHTDGINKHITTE	VNDVLK	FASYADLERITNWRRIAV	-YRANETKML	KTSYRCSNL	ANE--HFLELAVE	
TwTPS2	-TTRADRSKYFS	SIEVHDTL	KFPFANVERLAHRRNIEH	-HHVDNTRIL	KTSYCSNL	ANA--DFLOLAVE	
CfTPS14	GFILSRQLGRNINA	EVKQALE	SPFYAIMDRMAKRRSIEH	-YHIDNTRIL	KTSYCSN	FGNE--DFLSLSVE	
EpTPS8	-----LIENKSLYEE	VNFALGHP	-FANLDRLEARSYIDN	-YDPYDVPIL	KTSYRSSN	IDNK--DEWTIAFO	
EpTPS23	-----SISDKSLQNE	VDYVVKHP	-HAILRRLENRNYIEN	-YNTKNVSLR	KTSYRFSN	VNDKRS--DILAHSRQ	
SsSCS	-----SIPDKKLHKL	VEYLRNY	-KGITIRLGARNLEL	-YDMTYOALK	STNRFSNL	CNE--DFLVFAKQ	
AtEKS	-WVKTSVRD	KYLKKEVEDAL	AFPSYASLERSD	HRKILNGSA	VENTRVTKT	SYRLHNTCTS--DILKLAVD	
	500	510	520	530	540	550	560
CfTPS3	DFNISQAQHOKELQ	OLRWYAD	CRDLTLKFG	RDVVRIGNFLT	SAMIGDPEL	SDLRFAKHIVL	VTTRIDD
CfTPS4	DFNINQAQHRKELE	OLRWYAD	CRDLKLEF	GRDVRVSNFLT	SAILGDP	ELSEVRIVFAKHIVL	VTTRIDD
EpTPS1	DFNVQCSMHREELK	HILGRWVVE	KRLDKLKFARQK	IGYCYFSSA	SLFAPEMSD	ARI	SWAKNAVLT
TwTPS2	DFNRCQSIHREELK	HILGRWVET	KLDRLKFARQK	MAYCYFSA	AGTCFSP	ELSDARI	SWAKNSVL
CfTPS14	DFNRCQVIHREELR	ELERWVIEN	RDELKFA	RKSAYCYFSA	AATIFSP	ELSDARI	SWAKNGVL
EpTPS8	DFNKCQALHRVELD	YLEKWVKEY	KDLTLKWARQ	TEYALFTIG	AILSEPEY	ADARI	SWSONTVE
EpTPS23	DFNKCQIQFKELAY	LSRWEKXY	GLDKLYAROR	IEVVYFSIA	SNLFEPEF	SDARI	AWTOYAIL
SsSCS	DFDIHEAQNQKGLQ	OLRWYAD	CRDLTLNFG	RDVVI	IANYLASLI	TGDHAFDY	VRIFA
AtEKS	DFNFCQSIHREEMER	LDRWIVENR	LOELKFARQK	LAYCYFSG	AATLFS	PELSDARI	SWAKGGVL
	570	580	590	600	610	620	630
CfTPS3	FFDHGGPKKEESYE	LLELVKEWKEK	PAGEYVSE	EEVEILFTAV	YNTVNELA	EMAHIEQGR	SVKDLLV
CfTPS4	FFDHGGPREESHK	ILELKEWKEK	PAGEYVSE	KEVEILFTAV	YNTVNELA	ERANVEQGR	NVEPFLR
EpTPS1	FFDVGGSEELINL	VOLIERWDV	GSSEHCSEH	VEIVFSALH	STICEIG	EKAFAYQGR	RMRTSHV
TwTPS2	FFDIVGSEELANL	VHLLLENWD	ANGSPHYC	SEPVEIIF	SALRSTI	CEIGDKAL	AWQGRSV
CfTPS14	FFDVGGSVEELKN	LLOLVELWD	VDSRECI	SPSVQIIF	SALKHTRE	IGDKGFK	LOGRST
EpTPS8	FFDYGGSLECRNL	INIMHKWD	DHLTVGEL	SEKVEIVF	YSMYGT	INDLA	AKAEVRQ
EpTPS23	FFEYAASMDL	VNLINLIER	WDEHGSE	EFKSKEVE	ILFYALYD	LVNEDA	EAKAKYQ
SsSCS	FFDCHGSSQEC	KILELVKEW	KENPD	AEYGSSE	EILEFMAL	YNTVNELA	ERARVEQ
AtEKS	FFDVGGSKEELE	NLILHLEK	WDLVN	GPVEYSSE	HVEIIFSV	LRDTILET	GDKAFYQ
	640	650	660	670	680	690	700
CfTPS3	ILSVFRIELDTWT	NDTALTLE	EYLSQSW	SIGCRICIL	ISMOFQ	GVKLSDEM	LOSECTDI
CfTPS4	ILSIFKIELDTWS	DDTALTLD	DYLNNSW	SIGCRICIL	SMOFIG	MKLPEEMIL	SEECVDIL
EpTPS1	ILKSMILTETLW	SKSKATPT	LNEYMTNG	NTSFALGP	IVLPALF	FVGPKLTD	EDLKSHEL
TwTPS2	ILKSALREAEWA	RNVKVP	TFDEYVENG	YVSMALGP	IVLPALY	LIGPKVSE	EVVRSPE
CfTPS14	ILYSMMKESEW	GREKAVPT	IDEYISNAY	VSFALGP	IVLPALY	LIGPKLSE	EVVNHADY
EpTPS8	VMENMLKEREW	ADYNLVP	TFYEYVA	AGHTITGL	GPVILLAL	YFMGYPL	SEDDVQ
EpTPS23	ILKAMLKSESEY	VRYNIVPT	LDEYISNG	CTSSISFG	AILLIPLY	FTG--KMSE	EVVTSKE
SsSCS	ILSAFKIELDTW	SNGTQQSF	DEYISSSW	LSNGSRLT	GLETMOF	GVKLSDEM	LMSEECTDI
AtEKS	ILKSMILREAE	WSSDKSTP	SLEDYMEN	AYISFALGP	IVLPATYL	LIGPPLPE	KTVDSHOY

Fig. 4 continued

8 / 27

	710	720	730	740	750	760	770
CfTPS3	LLNDVQTEFEKERKENTG	NSVSLLOAAHKDERVIN	EEEA	CIKVKE	LAEYNRRKLMQIVYK	---	TGTLFP
CfTPS4	LLNDVQTEFEKERKENTG	NAVSLLLAAHKGERAF	SEEEA	IATAKAYLAD	CNRRSLMQIVYK	---	TGTLFP
EpTPS1	LLNDWRSYERESEEGKL	NAVSLHMIYGNCSVA	ATEEEA	TQIKGLIE	SERRELMRLVLQ	---	EKDSKIP
TwTPS2	LLNDTRTFKRESEAGKL	NSVLLHMIHSGS	---	GTTEEEA	VEKIRGMIADGR	RELLRLVLQ	---
CfTPS14	LLNDIRGMYERELKDGKL	NTLSLYMVNNEG	---	EISWEAA	ILEVKS	WIERERRELLRSVLE	---
EpTPS8	LLNDRVTIVKRESAOGKL	NCVSLFVEHGRG	---	AVDEETS	MKEVERLIVESHK	RELLRLIVOKTEGS	---
EpTPS23	LLNDRVTSQDMAOGKL	NSVSLRVLHSNG	---	TLTEEEA	KEEVDKILEK	HRRELLRMWVQ	---
SsSCS	LLNDVCSSEREREENIAGKSY	SILLATEKDGRKVSE	DEATAE	INEMVEYHWR	KVLQIVYK	---	KESTIP
AtEKS	LLNDIQGFRESAEGKL	NAVSLHMKHERD	---	NRSKEV	LTESMKGLAERK	REELHKLIVLE	---

	780	790	800	810	820	830
CfTPS3	CKDLEFLKACRIGCYLYSSG	DEFTSPQOMMEDMKS	LVYEPLP	ISPPEAN	NASGEK	MSCVSN
CfTPS4	CKDLEFLKVCRIKCYLYASG	DEFTSPQOMMEDMKS	LVYEPLQ	IHP	PAAA	---
EpTPS1	CKDLEFKMLKVLHMFYKDDG	FTS	---	NOMMKTAN	SLINOPT	ISHER
TwTPS2	CKDLEFKMVCVLHFLFYMDG	DGFTS	---	SPDMM	NAVNALIREP	ISL
CfTPS14	CKLELFWHMCTVHLFYSKDDG	FTS	---	QDLLSAVNAI	IYQPLVLE	---
EpTPS8	CKDLAWRVSKVLHLLYMDDG	FTCPVKMLNATNAIVNEP	LLITS	---	---	---
EpTPS23	CKKLEFWMTSKELHFLFYMTED	CFTCP	TKLLSAVNSTLKDPILMP	---	---	---
SsSCS	CKDVLEMAKGTFFYAVGINDEL	TSPQOSKEDMKSEVF	---	---	---	---
AtEKS	CKEATLMSKVINLEFYRKDDG	FTS	---	NDLMSLVKSVIYEPVSI	QKESLT	---

Fig. 4 continued

9 / 27

	10	20	30	40	50	60	70
CfTPS2
CfTPS1
TwTPS7	MHSLLMKKVIMYSSQTHVFPSP	LHCTIPKSSSFFLDAPVVR	LHC	ISGHGAKKKRLHFDIQGR	NATS		
TwTPS14/28
TwTPS21
EpTPS7
ZmAN2
syn-CPP
AtCPS
	80	90	100	110	120	130	140
CfTPS2	KQ	RSQVTECRVASLDAINGTQ	KVG	PATIGTPEEENKKT	DSIEYVKELTK	MGDGRISVSP	
CfTPS1	KK	ISKVAEIRVATVNAPP	VHDQD	DSTENQCHDAVN	NEDP	IEYIRTLRTITGDGRISVSP	
TwTPS7	KT	HTPEDLY-AKQEYSVPEI	IVK	DDDKKEE	VVKIKHEVDI	IKSMSSMEDGEISISA	
TwTPS14/28	KSKPARTTPEYS-DVLQ	TG	LPLIVE	DDIQEQEPELEV	SLENQIRQGV	DIVKSMIGSMEDGETSISA	
TwTPS21	KSKPARTTPEYS-DVLQ	TG	LPLIVE	DDIQEQEPELEV	SLENQIRQGV	DIVKSMIGSMEDGETSISA	
EpTPS7	NP	RIQEQTD-VFQKNGLP	VIKWHFVET	IDIDHEQVSKVS	VSNEIKKRVES	IKAIIDESMEDGDI	ITISA
ZmAN2	RA	QHTSESAAVAKGSS	LPIVRT	DAESRRTRWP	TDDDDAEPL	VD	IRAMLTSM
syn-CPP	RR	RRPCGMLISKSP	YPASEETR	EWEAEQHEHT	DELRETTT	MTIDGIRTA	LRSTIGE
AtCPS	KL	RTQEYINSQEV	QHDLP	LTHEWQ	QLQGEDAP	QISVGSNS	NAFKEAVKSVKTTIRN
	150	160	170	180	190	200	210
CfTPS2	YDTAIVALIKDL	EGGDP	EPFSSCLEWIA	QNLADGS	WGD	HFFCIYDRV	VNTAACVVALKSWNVHADKTE
CfTPS1	YDTAWVALIKDL	QGRDAP	EPFSSLEWIA	QNLADGS	WGD	AKFFCVMDRL	VNTIACVVALRSDVHAEKVE
TwTPS7	YDTAWVALIQDI	HNNGAP	QFPSSLLWIA	QNLDPGS	WGD	SRVFLAFDR	INTLACVVALKSWNVHPDKCE
TwTPS14/28	YDTAWVALVENI	HHPGSP	QFPSSLOWIA	QNLDPGS	WGD	PDVFLA	DR
TwTPS21	YDTAWVALVENI	HHPGSP	QFPSSLOWIA	QNLDPGS	WGD	PDVFLA	DR
EpTPS7	YDTAWVALVEDI	NGSGAP	QFPASLOWIA	QNLDPGS	WGD	AEIFTA	DR
ZmAN2	YDTAWVALVPR	LDGEGEP	QFPAAVR	WIRN	QNLDPGS	WGD	AAALFSAVDRLINTLACVVALTRWSLEPEMRG
syn-CPP	YDTSLVALIKRL	DGGDGP	QFPSTIDW	IQNLDPGS	WGD	ASFFMMGDR	IMSTLACVVALKSWNIHTDKCE
AtCPS	YDTAWVALID	AGDKTAP	FP	SAVKWIA	QNLSDGS	WGD	DAYLF
	220	230	240	250	260	270	280
CfTPS2	KGAVYLKENVHK	LKDGI	EHMPAGFE	VVPATLERAK	ALGIKGLPYDD	EFIREIYS	AKQTRITKIPKMI
CfTPS1	RGVRYINENVE	KLRDGN	EEHMTGCF	EVVFPALLQ	RAKSLGIOD	LPYDAPVIO	EIYHSREOKSKRIPLEM
TwTPS7	RGISFLKENIS	MLEKDD	SEHMLVGFE	FGFPVLLD	MARRLGI	DVPDDSP	FLQEIYVORDLKLKTRIPKDI
TwTPS14/28	RGLSFVKENIS	KLEKENE	EHMLIGFE	FAFPSSLE	MAKKLGI	EIPDDSP	ALQDIYTKRDLKLTRIPKDI
TwTPS21	RGLSFVKENIS	KLEKENE	EHMLIGFE	FAFPSSLE	MAKKLGI	EIPDDSP	ALQDIYTKRDLKLTRIPKDI
EpTPS7	RGMKYFREN	LCKLED	ENIEHMP	IGFEVAF	FPSSLE	ELAKLEI	QVPEDSPVLKDVYDSRNKLKTRIPKDI
ZmAN2	RGLSFLGRN	MWKLATE	DEESMP	IGFEVAF	FPSSLE	ELAKSLGVHDF	PYDHOALOGIYSSREIKMKRIPKVM
syn-CPP	RGLLFIOEN	MWRLA	HEEEDWML	VGFEVAF	FPSSLE	MAKOLDL	DIPYDEEALKAIIYAEERKLAKIPRDV
AtCPS	KGITTEFREN	ICKLED	ENIEHMP	IGFEVAF	FPSSLE	ETARGINI	DVPYDSEVLKDIYAKKELKLTRIPKDI
	290	300	310	320	330	340	350
CfTPS2	YESPTSLLYS	LDGLEGL	EWKILKLQ	SADGSF	ITSVSSTAE	VFMHTND	LKCHAFIKNALTNCNGGVPHTY
CfTPS1	HKVPTSLLS	SLEGL	ENLEWDK	LLKLOS	ADGSFLT	SPSSATA	AFMOTRDPKCYOFIKNTIOTFNGGAPHTY
TwTPS7	HNAPTTLHS	LEAIPD	LDWKLLK	LOCDGS	LLFSPSS	TAMAFINT	QENCLRYLYVYVORFNGGAPTVY
TwTPS14/28	HNVPPTLLS	SLEGL	PSLDWEK	LVKLOCT	DGSFLF	SPSSACAL	MHTKDCNCF
TwTPS21	HNVPPTLLS	SLEGL	PSLDWEK	LVKLOF	ONGSFL	SPSSATA	AFMHTKDCNCLSYLNDLVHKFNGGVPTAY
EpTPS7	HKVPTTLHS	SLEGM	PGLWEK	LLKLOS	DGSFLF	SPSSATA	YALMOTKQONCLEYLTIVHKFNGGVNVY
ZmAN2	HTVPTSLHS	SLEGM	PGLWAK	LLKLOS	DGSFLF	SPATAYAL	MNTGDDRCFSYIDRTVKKFNGGVNVY
syn-CPP	HAMPTTLHS	SLEGM	VDLDWEK	LLKLC	DGSFHC	SPSSATA	AFQOTGDKCFEYLDGIVKKFNGGVPCY
AtCPS	HKVPTTLHS	SLEGM	RDLDWEK	LLKLOS	DGSFLF	SPSSATA	AFMOTRQNCLEYLRNAVYKFNNGGVNVY

Fig. 5

SUBSTITUTE SHEET (RULE 26)

10 / 27

	360	370	380	390	400	410	420
CfTPS2	PVDLFEARLWAVDRLQRLGISRFFFEPEIKYLM	DHINNVRKGVFSSRHSOFADIDDTSMGIRLLKMHGYN					
CfTPS1	PVDVFGRIWADRLQRLGISRFFFESEIADCLAHIRHREWTEKCVTSGRESEFCIDIDDTSMGVRLRMHGYD						
TwTPS7	PVDLFEHNNWAVDRLQRLGISRFFFOPEIRECMSYVRYWTKDGICTNRNRVHDVDDTAMGFRLLRLHGYE						
TwTPS14/28	PVDLFEHIWCVDRQLQRLGISRFFHPEIKECLGYVHRYWTKDGCICWARNRSRVODIDDTAMGFRLLRLHGYE						
TwTPS21	PVDLFEHIWSDRLQRLGISRFFHPEIKECLGYVHRYWTKDGCICWARNRSRVODIDDTAMGFRLLRLHGYE						
EpTPS7	PVDLFEHIWAVDRLQRLGISRFFOPOLKOSVDYVARYWEEDGICWARNSSVHDVDDTAMGFRVLRSGFHH						
ZmAN2	PVDLFEHIWAVDRLRLGISRFFFOKEIEECMDYVNRHWTEDGICWARNSDVKEVDDTAMFRLLRLHGYE						
syn-CPP	PIDVYERLWAVDRLRLGISRHFTSEIEDCLDYIFRNWTEDEGIAHTKNCVVKDIDDTAMGFRLLRLHGYE						
AtCPS	PVDLFEHIWIVDRLQRLGISRFFFEIKECLDYVHRYWTKDGCICWARNSHVODIDDTAMFRLLRLHGYE						
	430	440	450	460	470	480	490
CfTPS2	VNPNALEHFKOKGKFTCYADQHI-ESPSPMYNLYRAQLRFPGEE-ILQALQFAYNELHENLASNHFO						
CfTPS1	VDPNVLKNEK-KDDKFSYGGQMI-ESPSPITNLYRASQLRFPGEE-ILEDANKFAYDFLOEKLAHNQIL						
TwTPS7	VHPDAFRQEK-KGCEFIQYEGQSH-PTVTVMYNYLRASQLMFPEEK-ILDEAKQFTEKFLGEKRSANKLL						
TwTPS14/28	VSPDVFKQER-KGDEFVCFMGQSN-QAATGCHYNLYRASQMMFPEET-ILEEAKKFSVNFIREKRAASELL						
TwTPS21	VSPDVFKQER-KGDEFVCFMGQSN-QAATGCHYNLYRASQMMFPEET-ILEEAKKFSVNFIREKRAASELL						
EpTPS7	VSADVFKEBK-KGDTFFCFAGQST-QAVTGMYNLIRASQLMFPEEK-ILEEAKQFSSAFLKVKODANEVL						
ZmAN2	VSPDVFKNEE-KGGEFFAFVGQSN-QAVTGMYNLNRASQLSFPGED-VLHRAGAFSYEFIRKKEAEGALR						
syn-CPP	VDECVLKKEE-KDGKFFCLHGESNPSSVTMYNTYRASQLKFPDGDGVLGRAEVECRSELODRRGSNRMK						
AtCPS	VSADVFKNFE-KEGEFFCFVGQSN-QAVTGMENLYRASQLAFPREE-ILKNAKKFSYNVLEKREEREELT						
	500	510	520	530	540	550	560
CfTPS2	EKWVISDHLIDEVRIGLKPWPYATLPRVEASYYLQHYGGSSDVWIGKTLYRMPPEISNDTYKILAOOLDENK						
CfTPS1	DKWVISKHLDEIKLGLMPWYATLPRVEARYIIOYVAGSGDVWIGKTLYRMPPEISNDTYHELAKTDKRR						
TwTPS7	DKWIIITKDLPGEVGFALDPWYASLPRVEARFEIIOHYGGEDDVWIDKALYRMPYVNNNVYLELAKLDYNY						
TwTPS14/28	DKWIIITKDLNEVGAFALDPWYACLPRVETRLYIEOYGGODDVWIGKTLYRMPYVNNNVYLELAKLDYNN						
TwTPS21	DKWIIITKDLNEVGAFALDPWYACLPRVETRLYIEOYGGODDVWIGKTLYRMPYVNNNVYLELAKLDYNN						
EpTPS7	DKWIIITKDLPGEVKYALDPWYASLPRVESREYIEOYGGSDDVWIGKTLYRMPYVNNNDYLLAKLDYNN						
ZmAN2	DKWIIISKDLPGEVVYTLDPWYCNLPRVEARDYIEOYGGSDDVWIGKTLYRMPYVNNNDYLLAKLDYNN						
syn-CPP	DKWAIITKDLPGEVVYADPWKASLPRVETRLYIIOYGGSDDVWIGKTLRMTLFCNDLYLKAADKTSN						
AtCPS	DKWIIITKDLPGELGFALEDPWYASLPRVETREYIIOYGGENDDVWIGKTLYRMPYVNNNGYLELAKODYNN						
	570	580	590	600	610	620	630
CfTPS2	COAQHOLEWMSMKWYQSNVKEFG-ISKKELLFLAYFLAAATMFEPERTQERIMWAKTQVVSMTSEL						
CfTPS1	COAQHOFEWIYMOEWYESCNMEEFG-ISRKELLVAYFLATASIFELERANERLAWAKSQIISTITASEF						
TwTPS7	COALHRTWGHIOKWEYCKPRDFG-ISRECLLRAYFMAAASIFEPEPERSMERLAWAKTATLLEIIVSMF						
TwTPS14/28	CQSLHRIEWDNIQKWEYGNLGGFG-VNKRSLRLTYFLATSNIFEPEPERSVERLTWAKTATLVOATASYE						
TwTPS21	CQSLHRIEWDNIQKWEYGNLGGFG-VNKRSLRLTYFLATSNIFEPEPERSVERLTWAKTATLVOATASYE						
EpTPS7	COAVHRSEWDNIQKWEESDLAEFG-VSRREILMAYFLAAASIFEPEKSRERLAWAKTSVLINTLOAYE						
ZmAN2	COALHOLEWQGLKRWYTNRLMDFG-VAQEDALRAYFLAAASVYEP CRAERLAWARAAILANAVSTHL						
syn-CPP	FOKECRVELNGLRRWYLRNLERFGTDPQTTLMTSYFLASANIFEPNRAAERLGLWARVALLADAVSSHE						
AtCPS	COAQHOLEWDIFOKWYENRLSEWG-VRRSELLECYYFLAAATIFESERSHERMVWAKSSVLVKAISSSE						
	640	650	660	670	680	690	700
CfTPS2	NKENTMSFDLKIALLTQPOH-----QINGSEMK-----NGIAQTLPAAFRQLLKEFDKYTR						
CfTPS1	NNQNTSPED-KLAFLTDFK-----NGNSTN-----MALVT-----LTQFLEGFDRTYS						
TwTPS7	NEVGNSTEQRIAFTEFSIRASPMGGYINGR-KLDKIGTTQE-ILIQMLATIDQFSDAFAAYGHQIT						
TwTPS14/28	E---NSREERIEFANEFQKFPNTRG-YINGR-RLDVKQATKG-LIEMVFATLNQFSLDALVVHGEDIT						
TwTPS21	E---NSREERIEFANEFQKFPNTRG-YINGR-RLDVKQATKG-LIEMVFATLNQFSLDALVVHGEDIT						
EpTPS7	HENNSTIHEKAAAFVQLEKS-----GFATNAR-KLEG-KTMEK-LGRLIVGTINDVSLDTAMAYGKQIS						
ZmAN2	RNSPSFRFRLHSLRCRPSSEETDGSWFNSSSGSDAVLVKAVLR-LTDSLAREAQPIHGDPEDITHKLL						
syn-CPP	RR-----ICGPKN-LTSNLEELISLVP-----FDDAYSGSL						
AtCPS	GESSDSRRSFSDQFHEYIANARRSDHFNDRNMRDRPGSVQASRIAGVILGTTNOMSFLLFMSHGRVNV						

Fig. 5 continued

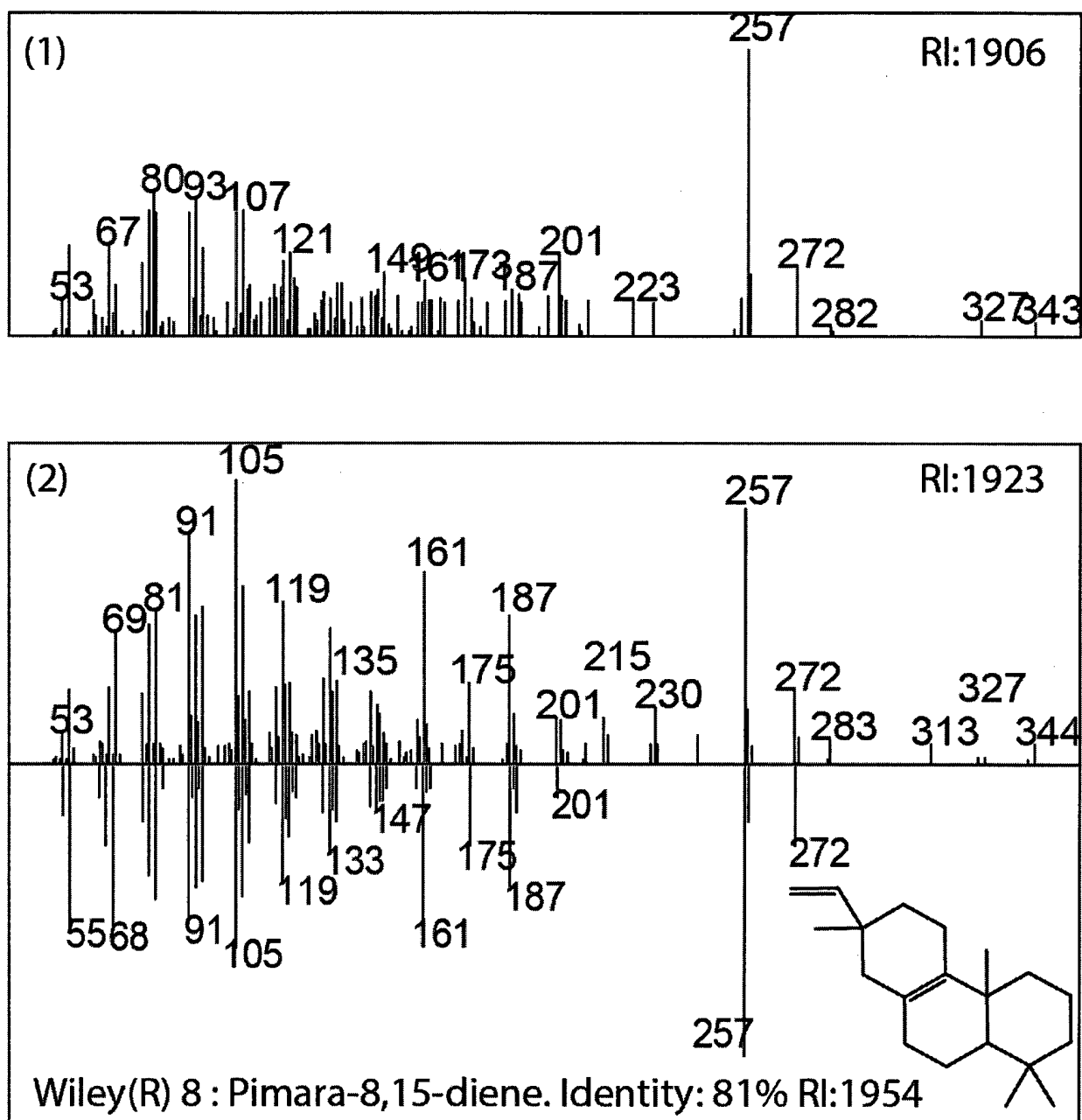
SUBSTITUTE SHEET (RULE 26)

11 / 27

	710	720	730	740	750	760	770
CfTPS2	HQLRNTWN	-----	KWLMKLGDDNGGADAELLANTLNICAGHN	-----	EDILSHYEYTAUSS		
CfTPS1	HQLKNWS	-----	VWLRKLQGGEGNGGADAELLVNTLNICAGHIAFREILAHNDYKTLN				
TwTPS7	RHLHNSWK	-----	MWLLKWQEEGDRWLGEAELLITOTINLMADHKIAEKLFMGHTNMEQLFS				
TwTPS14/28	HHLVQSWE	-----	KWVLTWQEGGDRREGAELLVOTINLMAGHTHSQE	-----	EELMERLFFK		
TwTPS21	HHLVQSWE	-----	KWVLTWQEGGDRREGAELLVOTINLMAGHTHSQE	-----	EELMERLFFK		
EpTPS7	RDLRHWD	-----	ICLOKWEESGDMHOGAOLLVNTINLTSDAWNFD	-----	LSSHMHQFFQ		
ZmAN2	RSAAAEVVRKADAADSVCNGSSAVEQEGSRMVHDKOTCLLIARMTEISAGR	-----	AAGEAASEDGDRRTIO				
syn-CPP	REAWKQWL	-----	MAWTAKESSQESIEGDTAILLVRAEILFGGRHVLTGQRPDLWEYSOLEQ				
AtCPS	NLLYLSWG	-----	DWMEKWKLYGD	-----	ECGELVVKMILIMKNNDLTNFFT	-----	HTHEVRLAE
	780	790	800	810	820	830	840
CfTPS2	LTNKKICRLSQIQDKMLEIE	-----	EGSIKDKEMELETOTLVKLVLE	-----	TSGGIDRNIKOT		
CfTPS1	LTSKICRLSQIQNEKELETEGQ	-----	KTSIKNKELEEDMORLVKLVLEK	-----	SRVGINRDMKKT		
TwTPS7	LTNKVCYSLGH-HELQNNKE	-----	LEHDMORLVOLVLN	-----	SSDGLDSIKKT		
TwTPS14/28	LTNTVCHQLGHYHHLNKKDKPQOVEDNGGYNNSNPESISKLOIESDMRELVLVLN	-----	SSDGMDSNIKOT				
TwTPS21	LTNTVCHQLGHYHHLNKKDKPQOVEDNGGYNNSNPESISKLOIESDMRELVLVLN	-----	SSDGMDSNIKOT				
EpTPS7	LVNEICYKIRKRYKKNKVNDK	-----	KKTTTPETESHMOELVKLVLE	-----	SSDDLDSNLKQI		
ZmAN2	LTGSIQDSIKQKMLVSQDPEKNEE	-----	MMSHVDDDELKLRIREFVOYLRLGEKKTGSSETROT				
syn-CPP	LTSSICRKLRYRRLAQEN-GKSTE	-----	KVEIDQQLDLEMOELTRRLVLOGCSAIN	-----	RLTRET		
AtCPS	IINRICLPRQYLKARRNDEK	-----	EKTIKSMEKEMGKMVELALS	-----	ESDTFR-DVSIT		
	850	860	870				
CfTPS2	FLSVFKLFYYRAYHDAKTIIDAHIFQVLFEPPV	-----					
CfTPS1	FLAVVKLFYYKAYHSAQAIDNHMFVLEFPVA	-----					
TwTPS7	FLAVAKRFYYTAFVDEETVNVHIAKVLFEFVD	-----					
TwTPS14/28	FLAVTKSFYYTAFTHPGTVNYHIAKVLFEFVV	-----					
TwTPS21	FLAVTKSFYYTAFTHPGTVNYHIAKVLFEFVV	-----					
EpTPS7	FLTVARSFYPAVCDAGTINYHIAKVLFEFVY	-----					
ZmAN2	FLSLVKSYYAAHCPPHVVDRIHSRVLEFPVSAAK						
syn-CPP	FLHVVKSECYVAYCSEETIDNHIDKVLFEQDVI	-----					
AtCPS	FLDVAKAFYYFALCG-DHLQTHISKVLFEQKV	-----					

Fig. 5 continued

12 / 27

**Fig. 6**

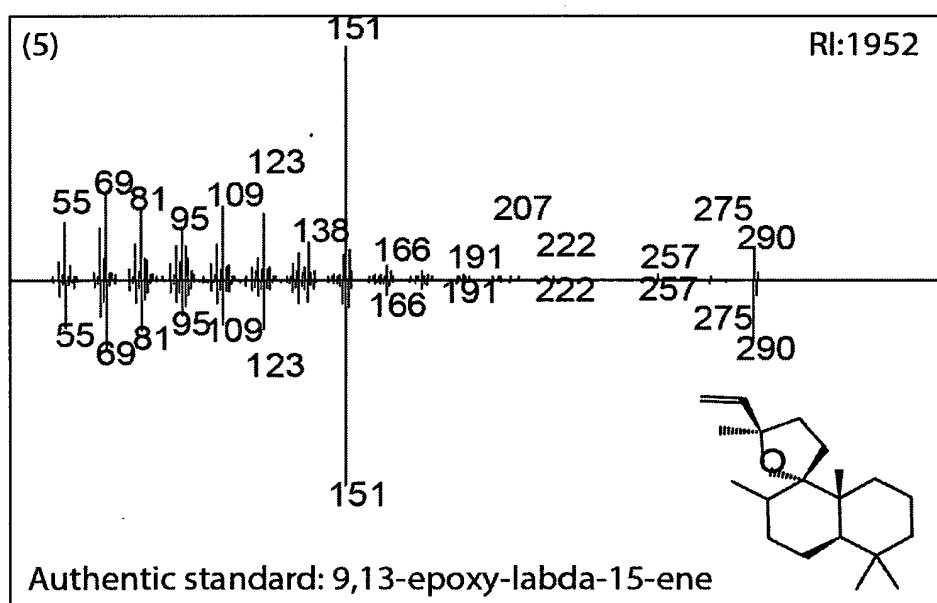
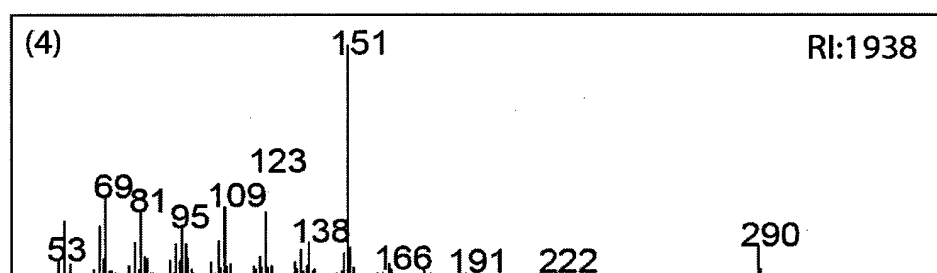
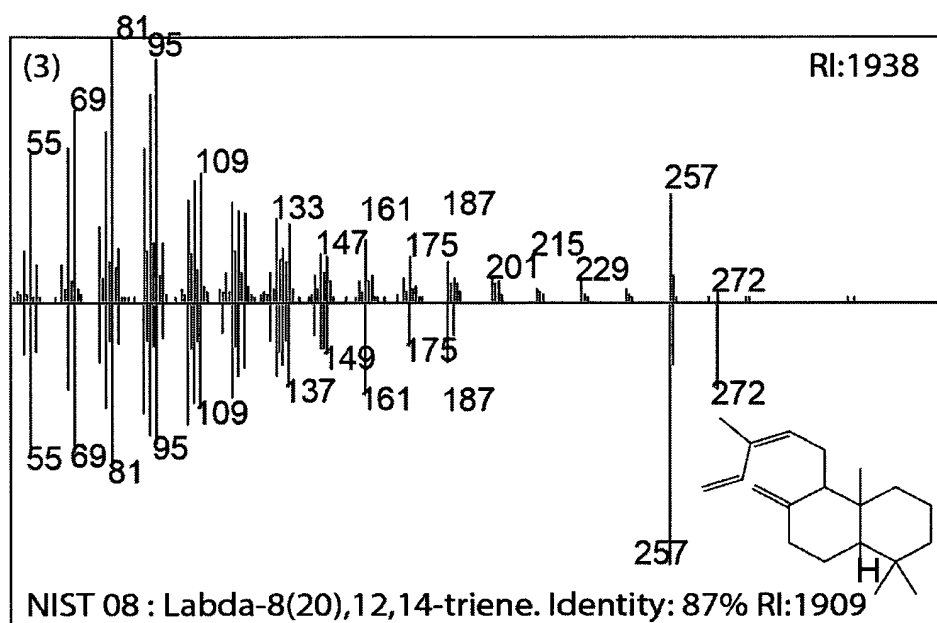
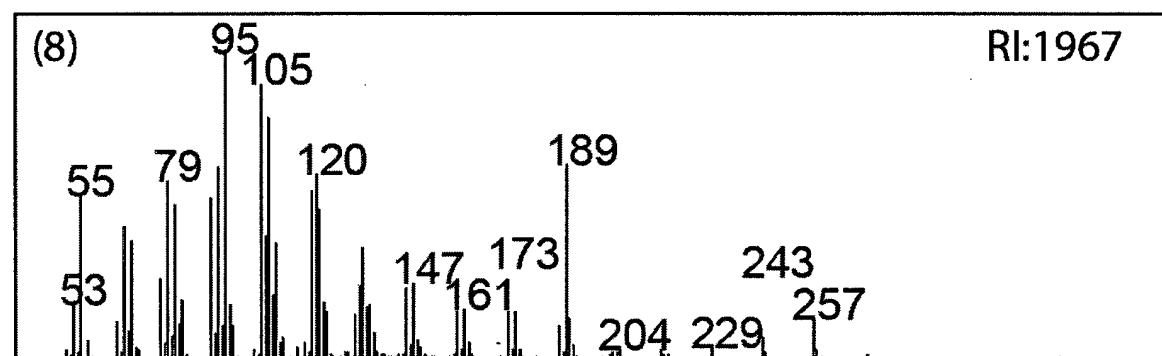
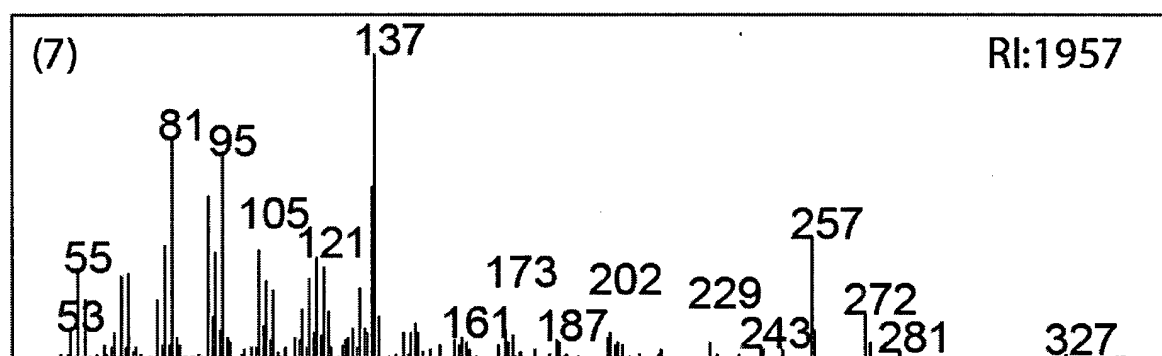
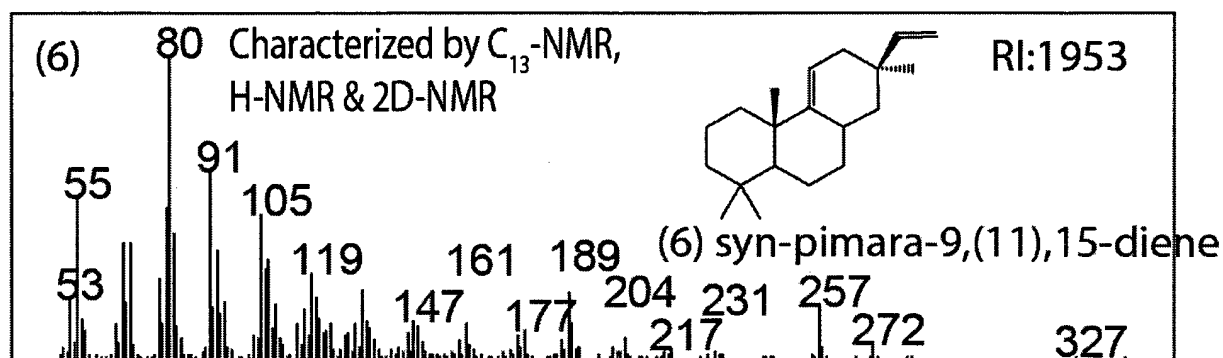
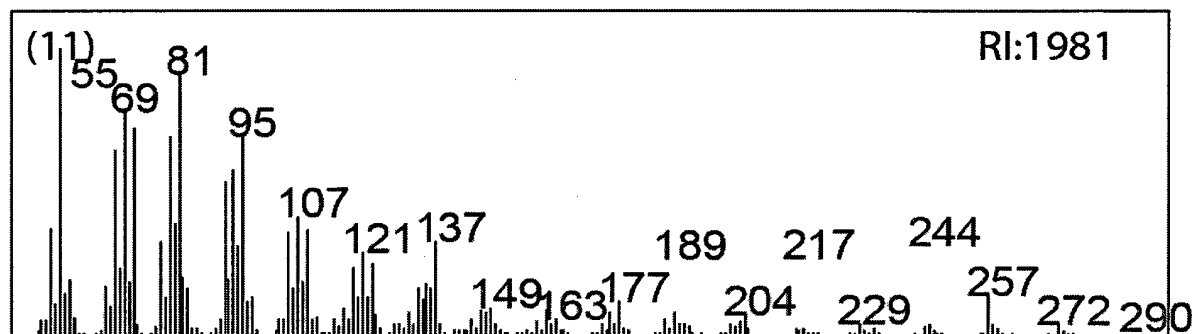
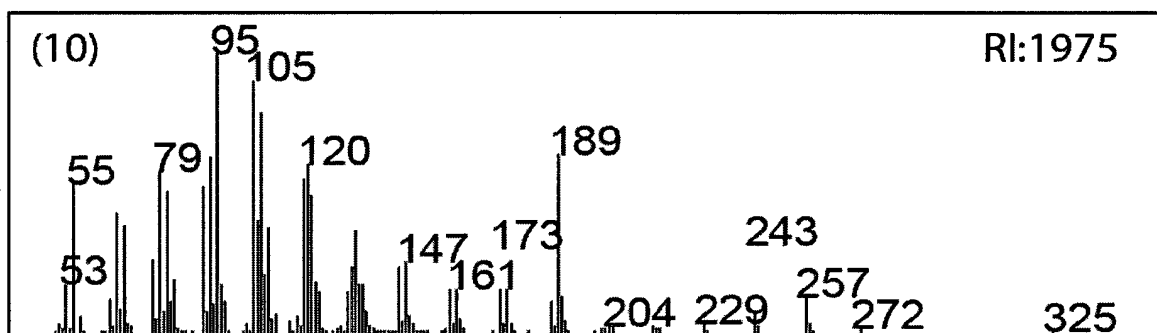
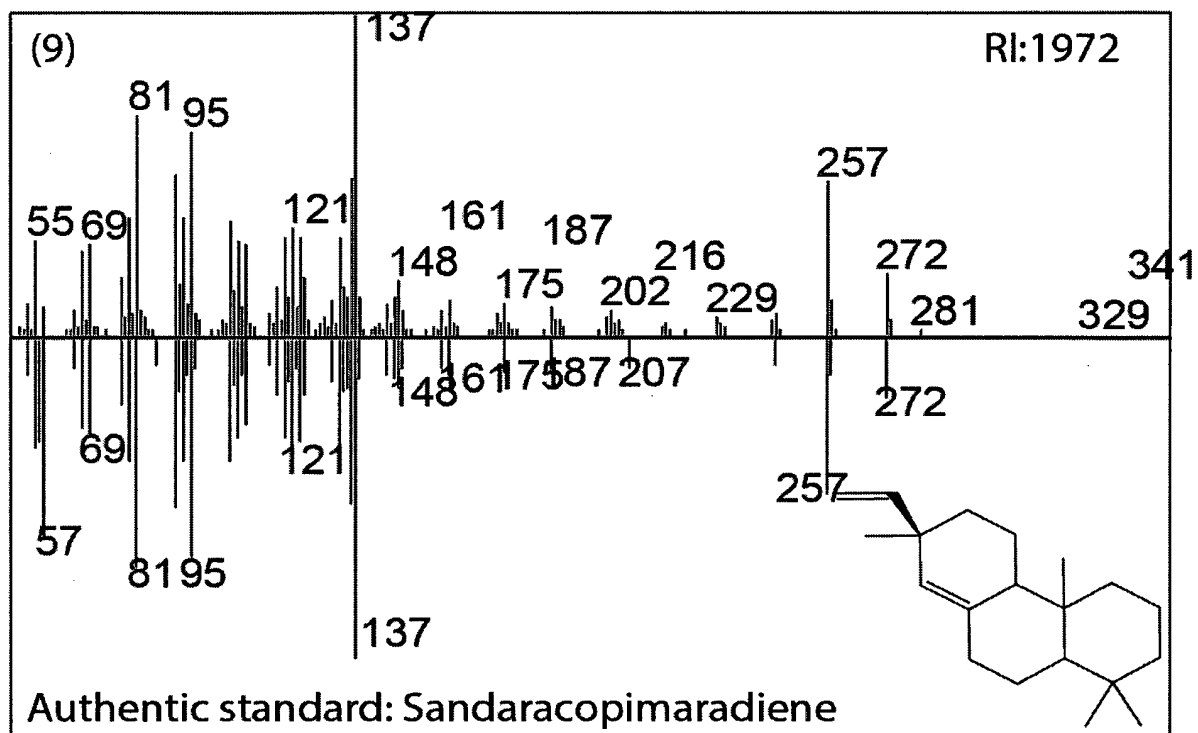
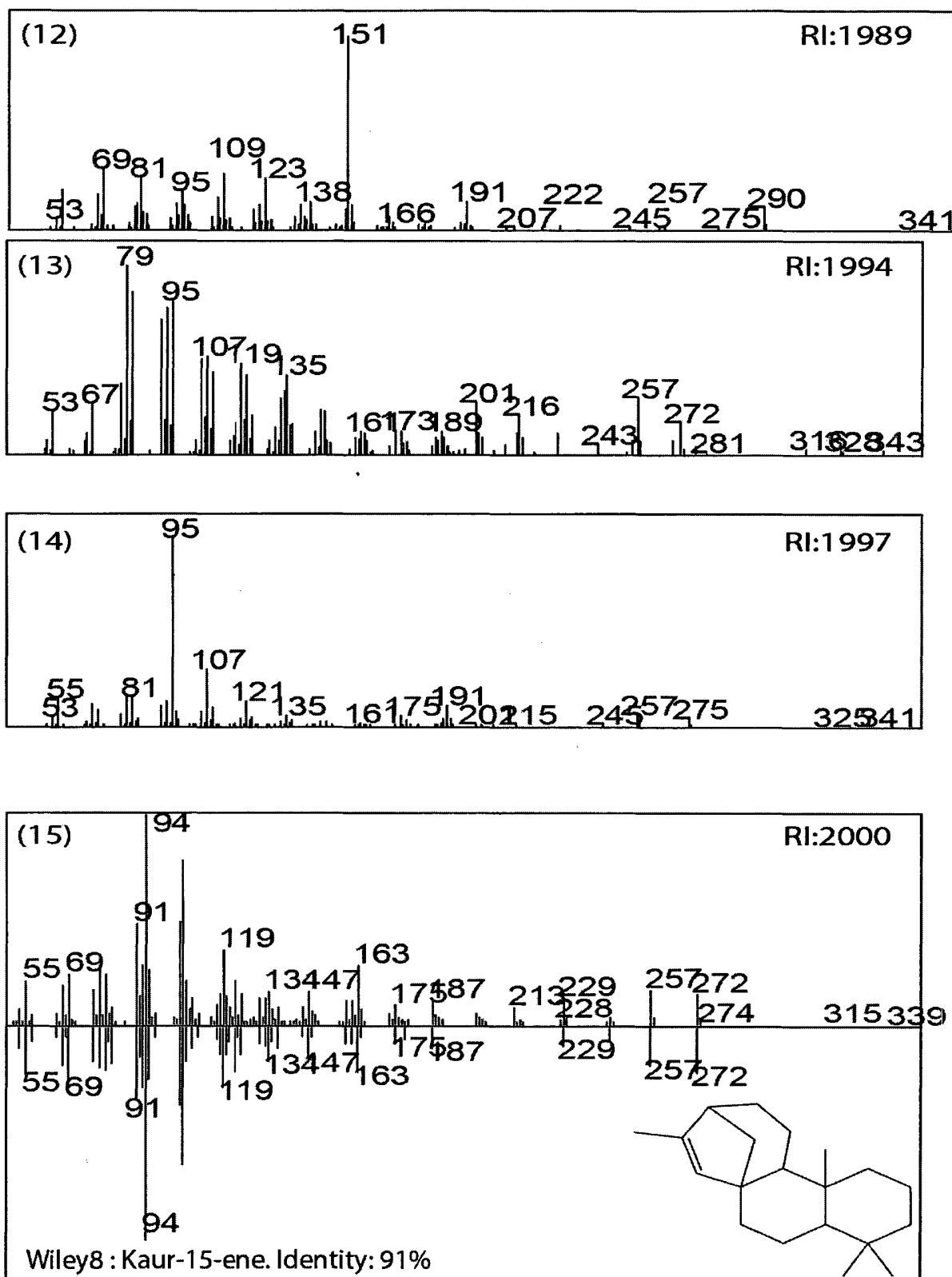


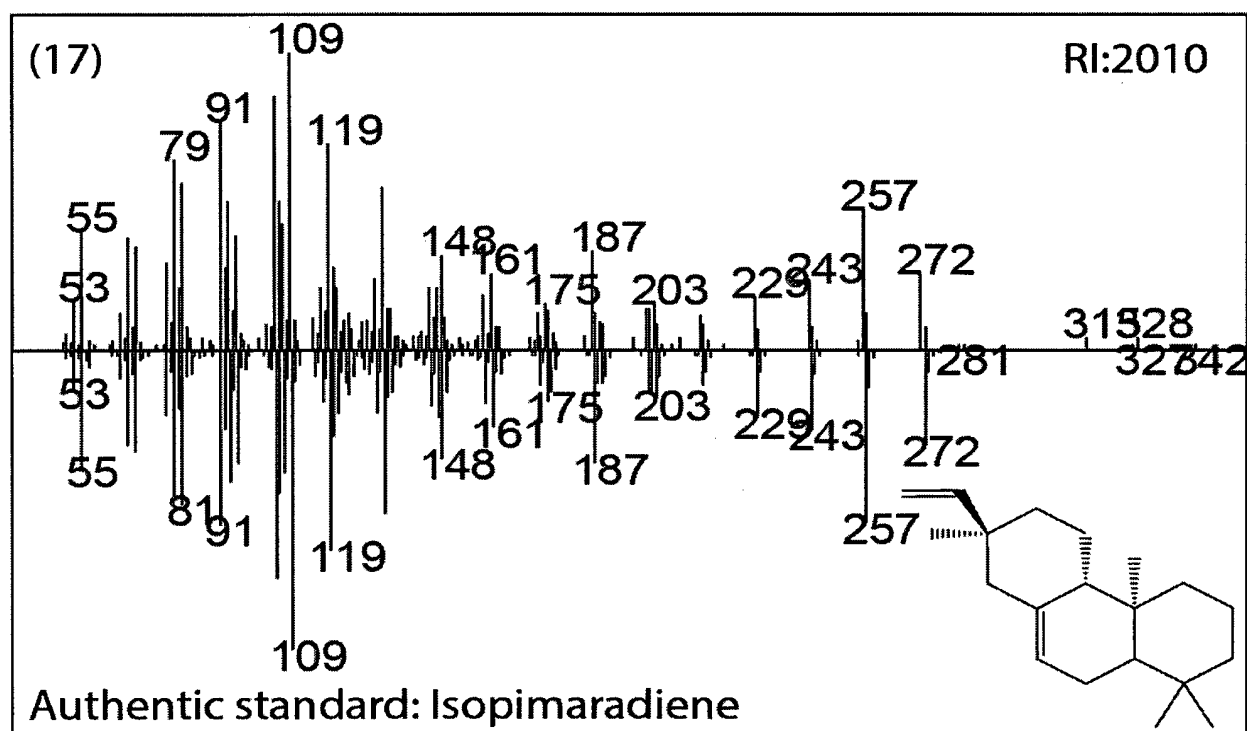
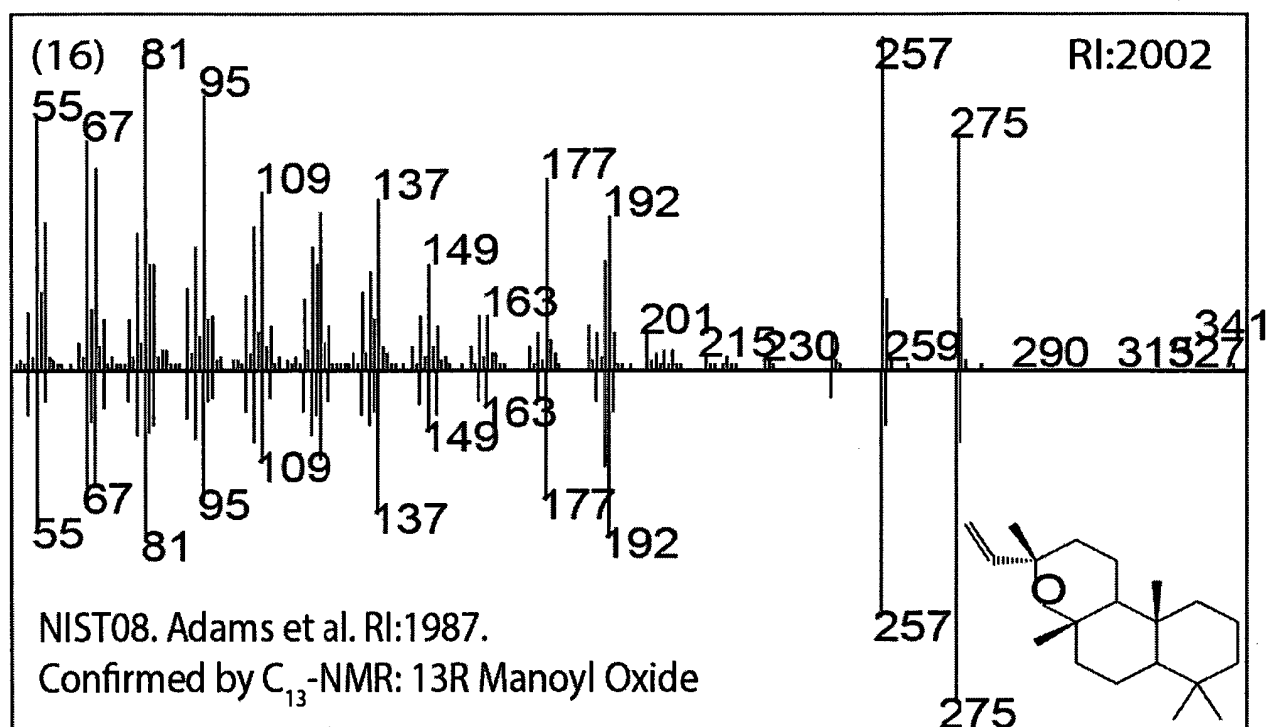
Fig. 6 continued

**Fig. 6 continued**

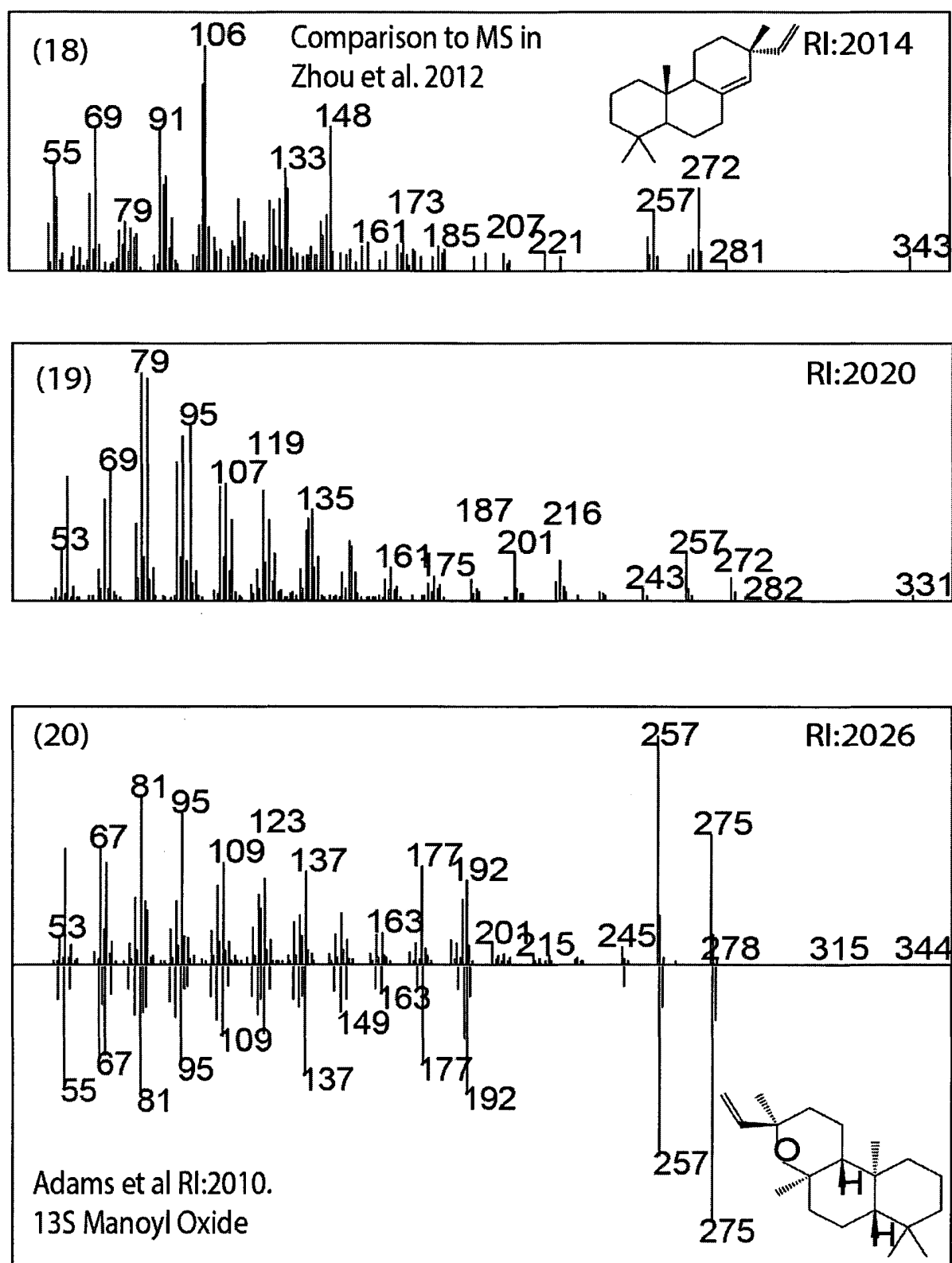
**Fig. 6 continued**

**Fig. 6 continued**

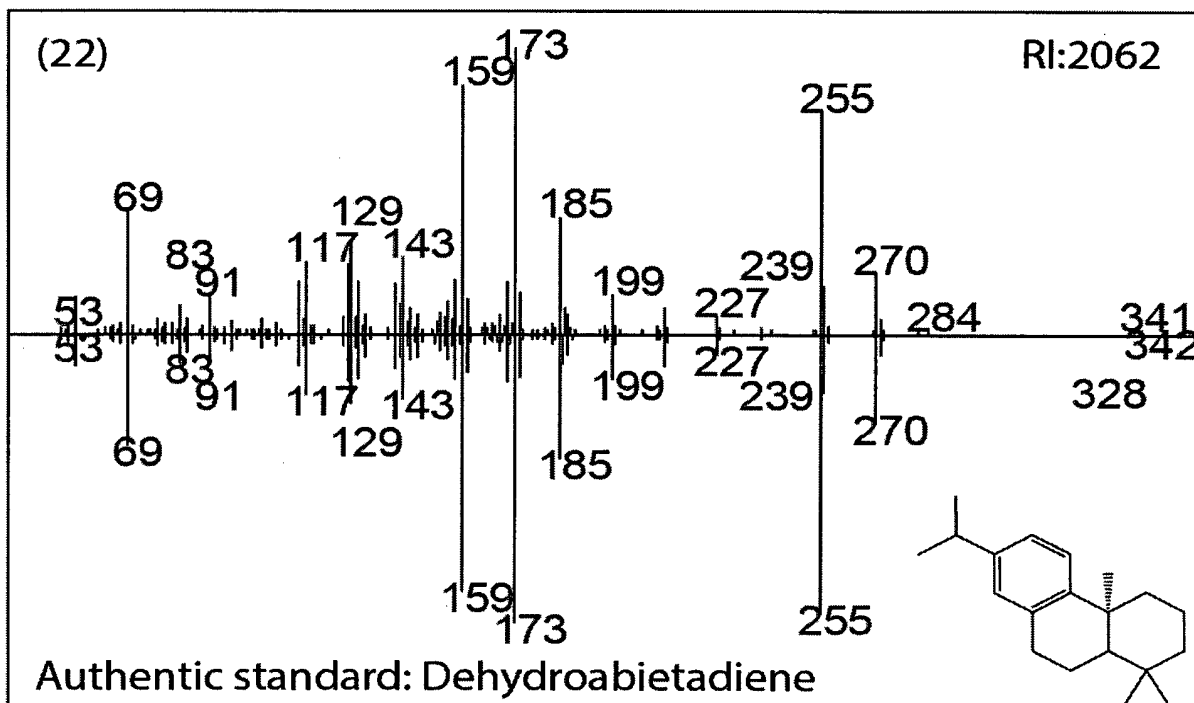
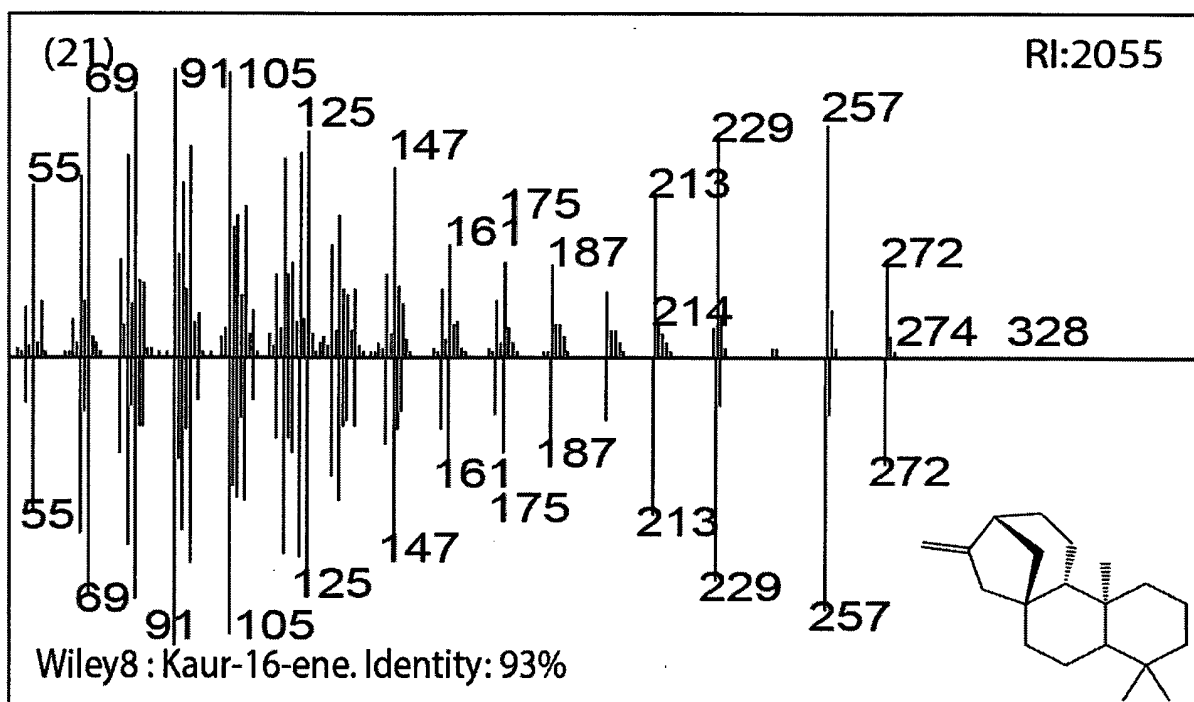
SUBSTITUTE SHEET (RULE 26)

**Fig. 6 continued**

18 / 27

**Fig. 6 continued**

SUBSTITUTE SHEET (RULE 26)

**Fig. 6 continued**

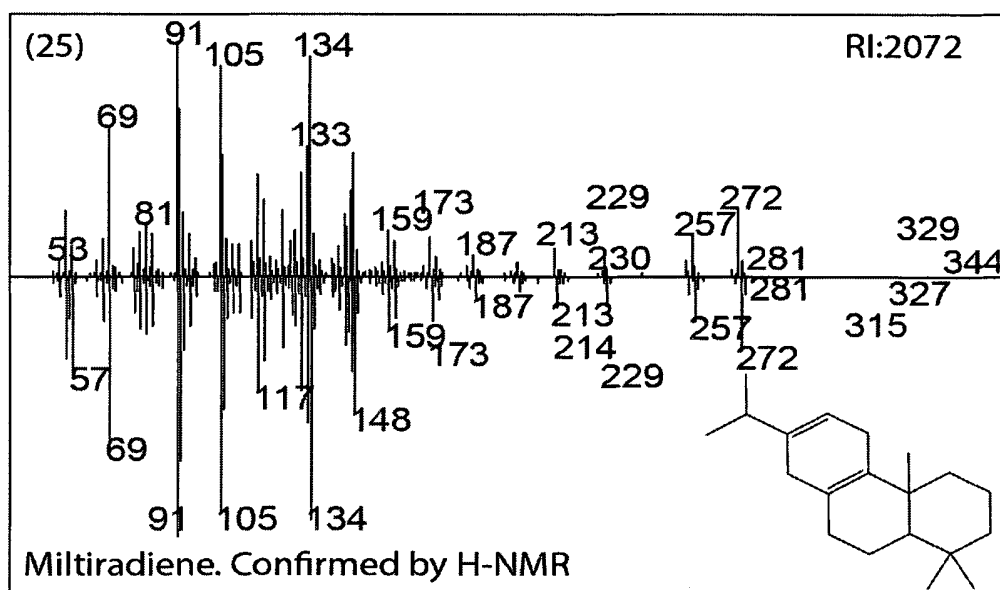
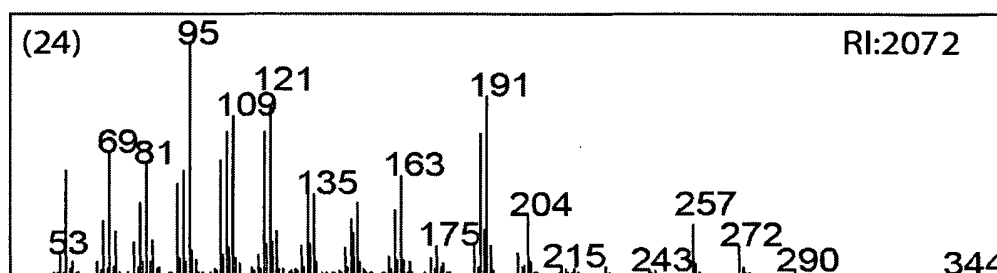
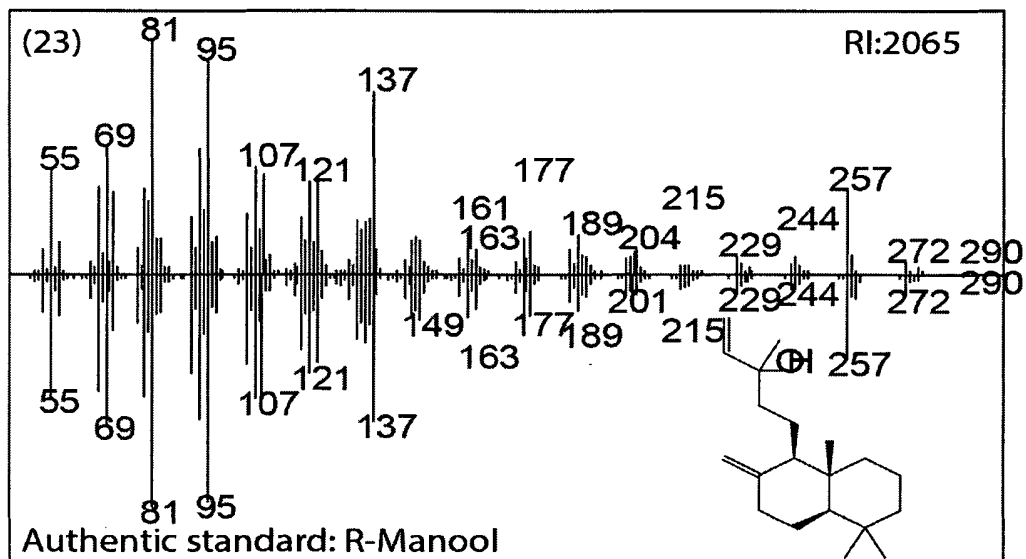
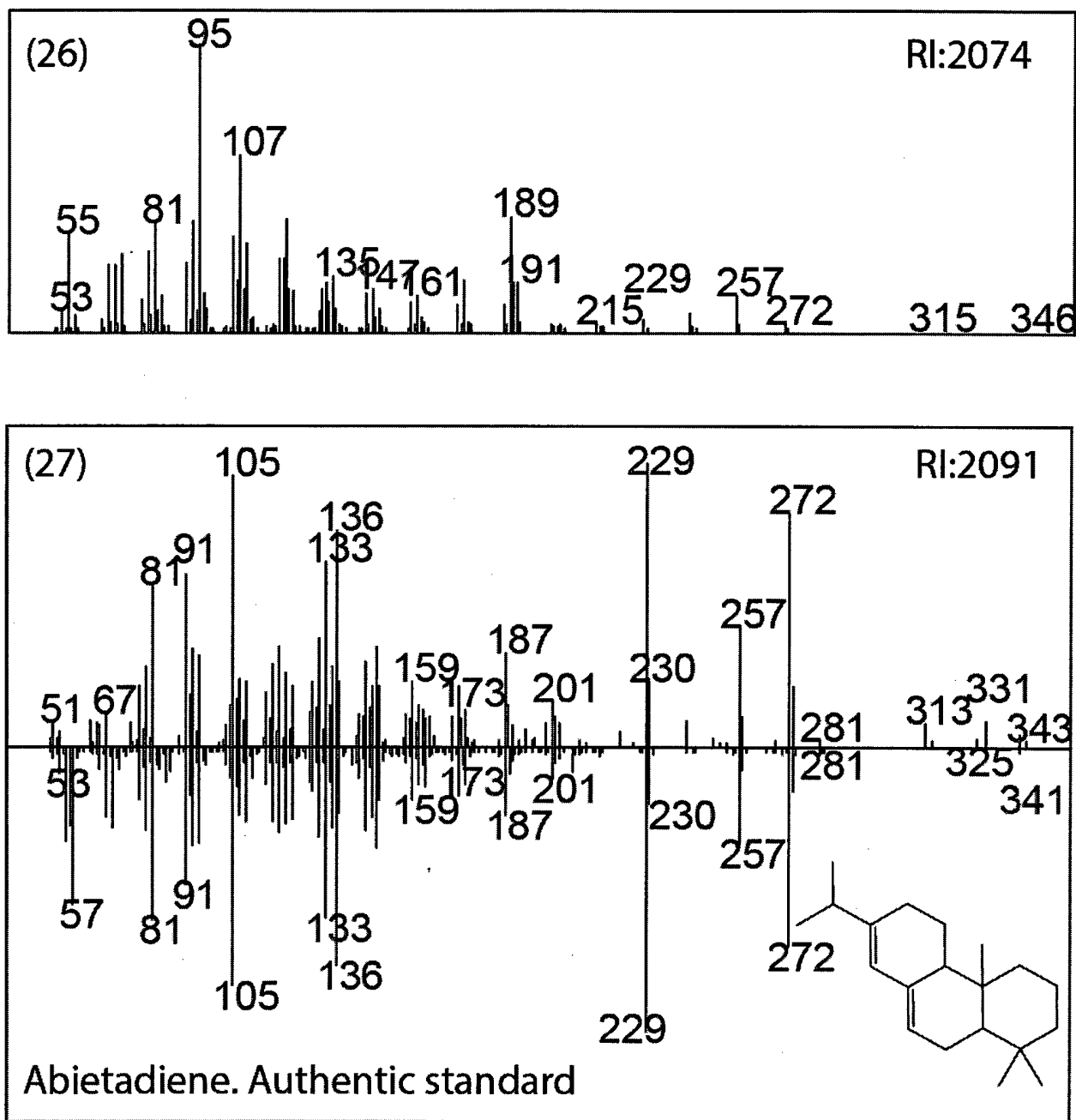
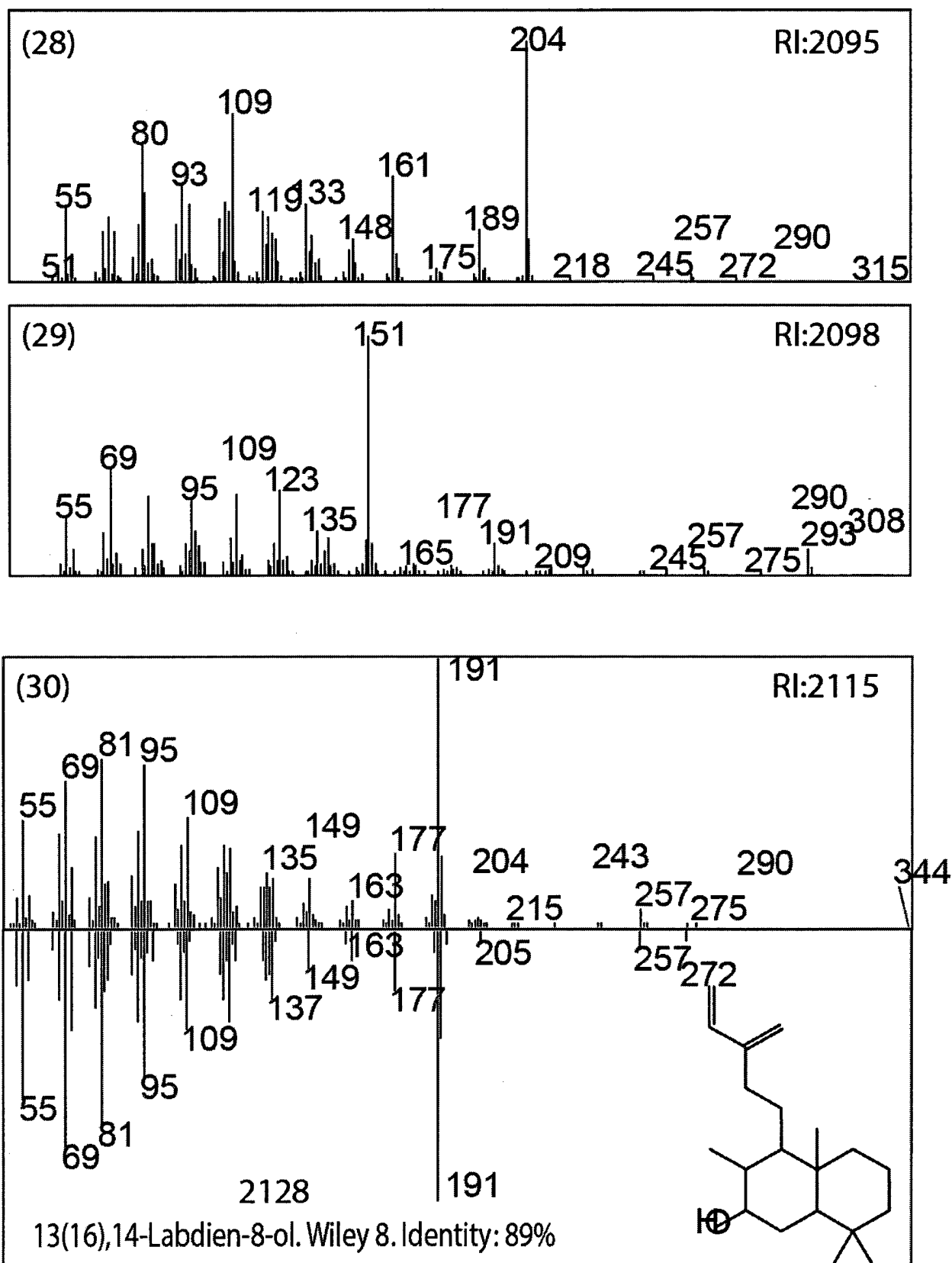
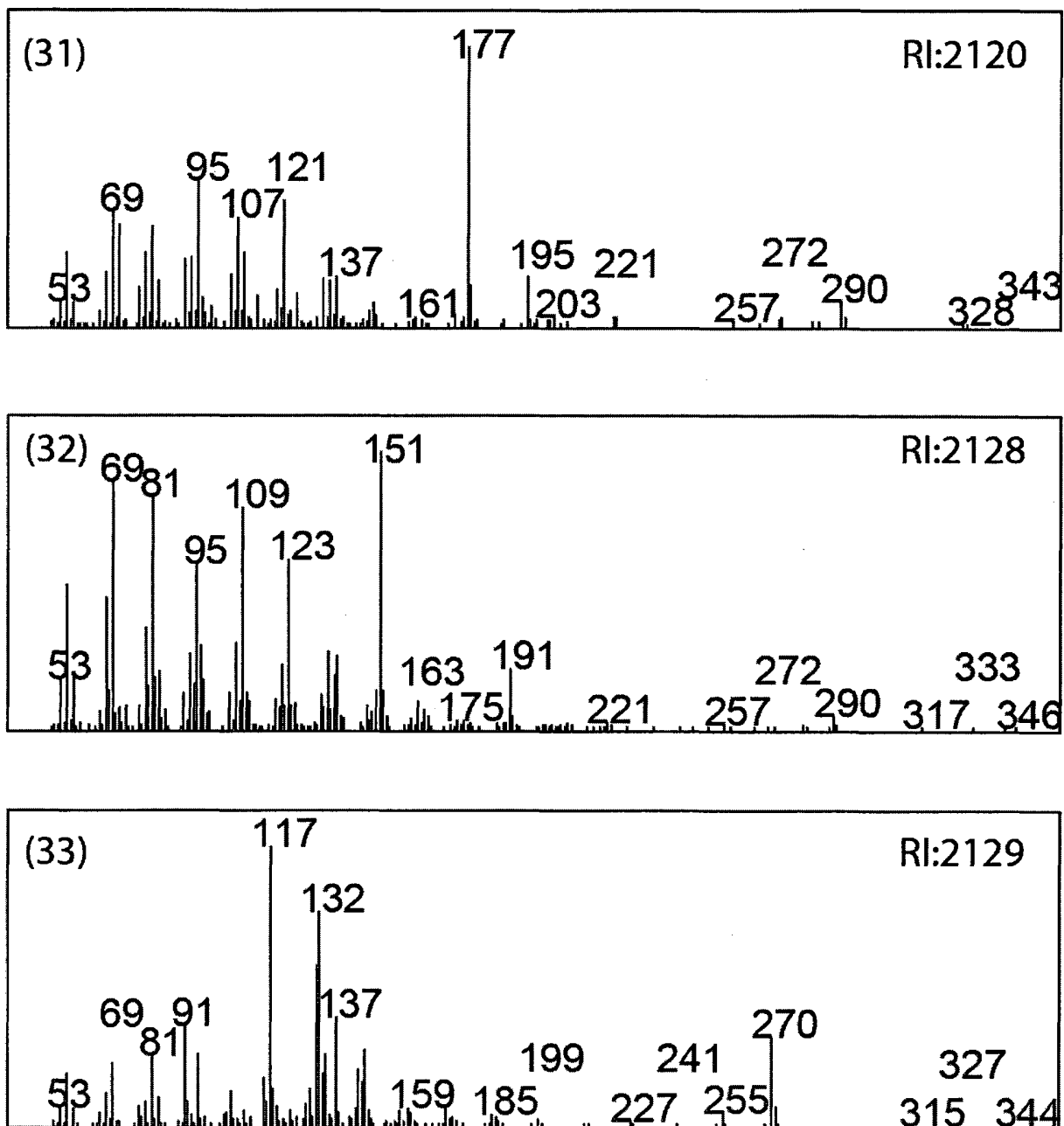


Fig. 6 continued

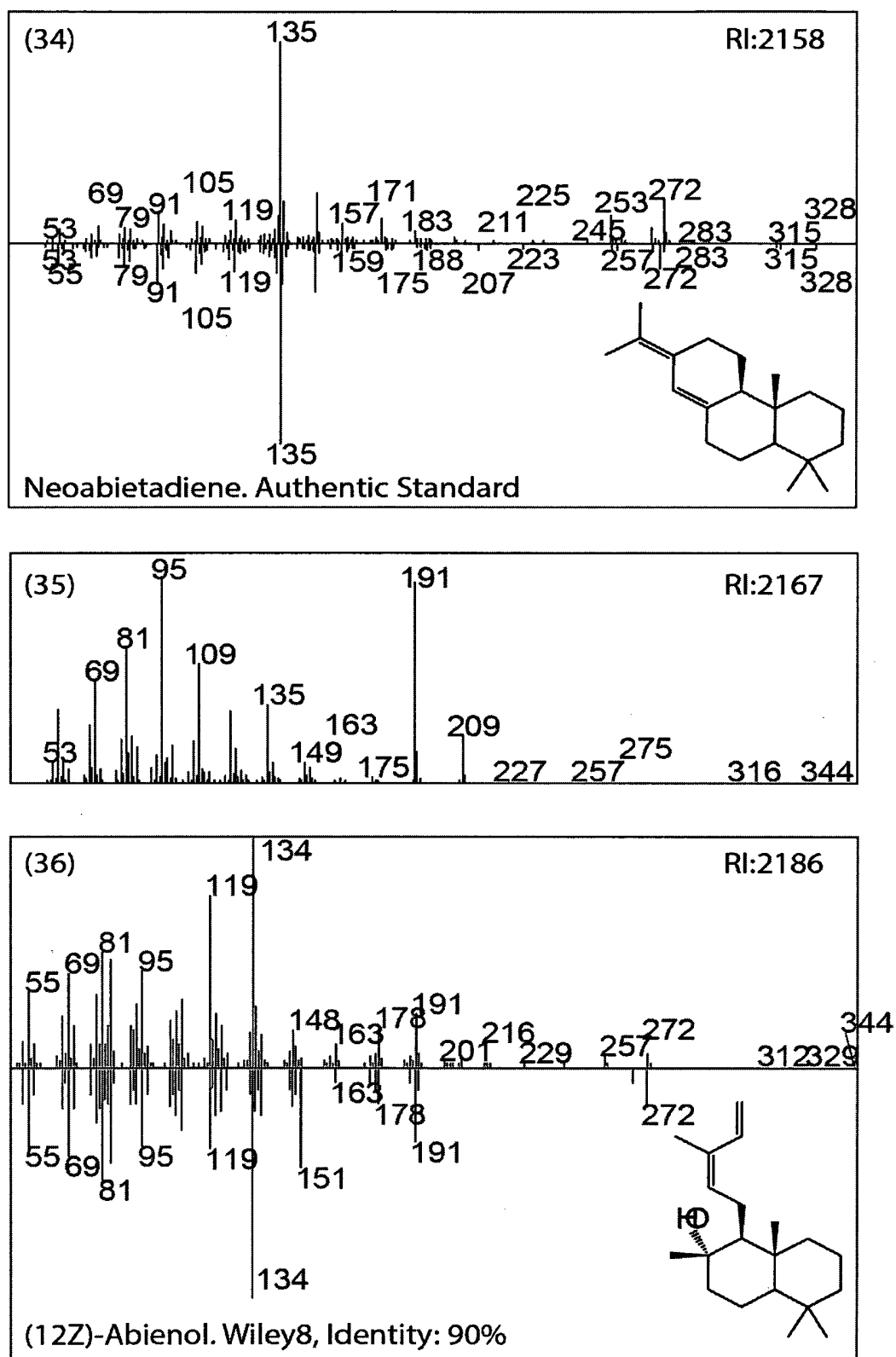
**Fig. 6 continued**

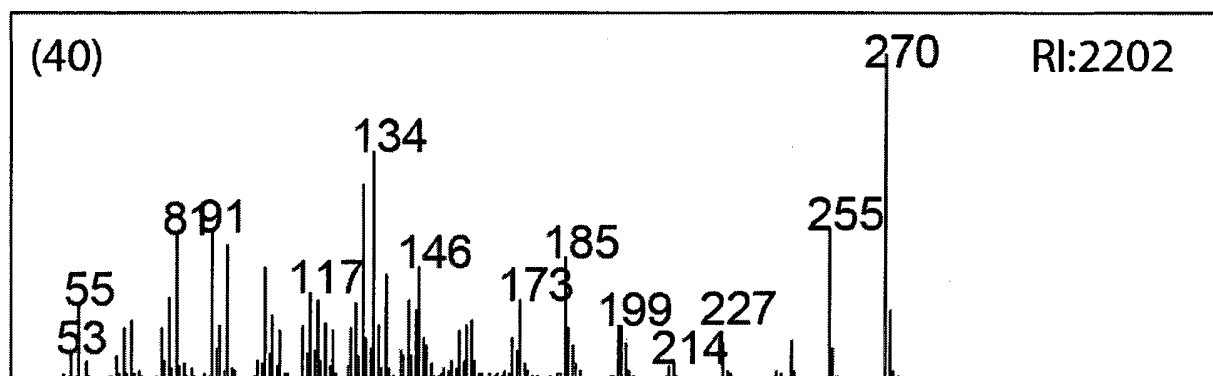
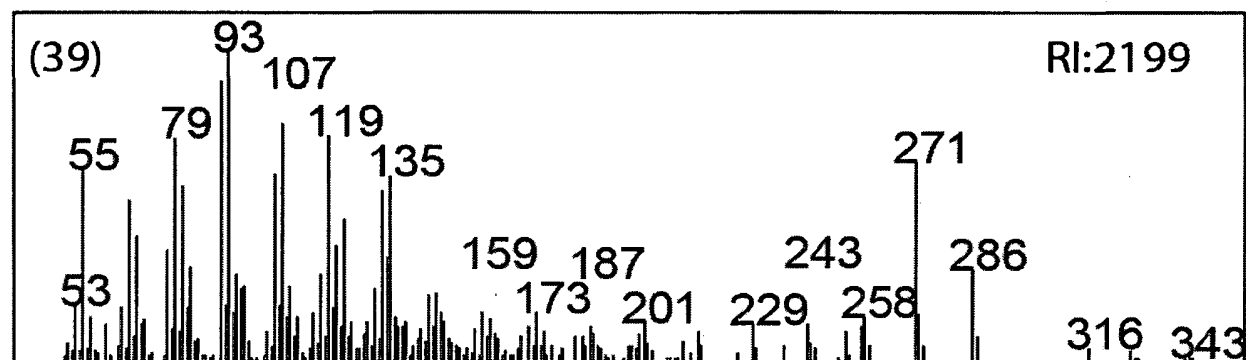
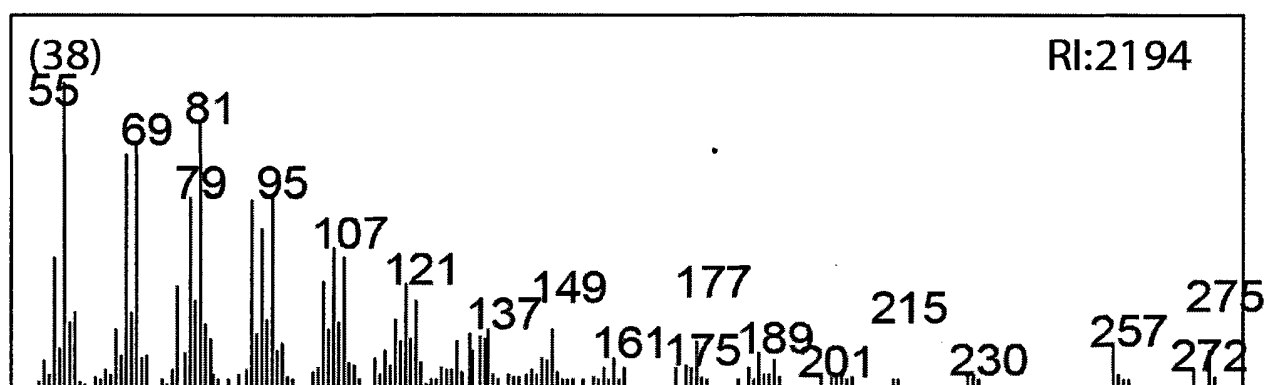
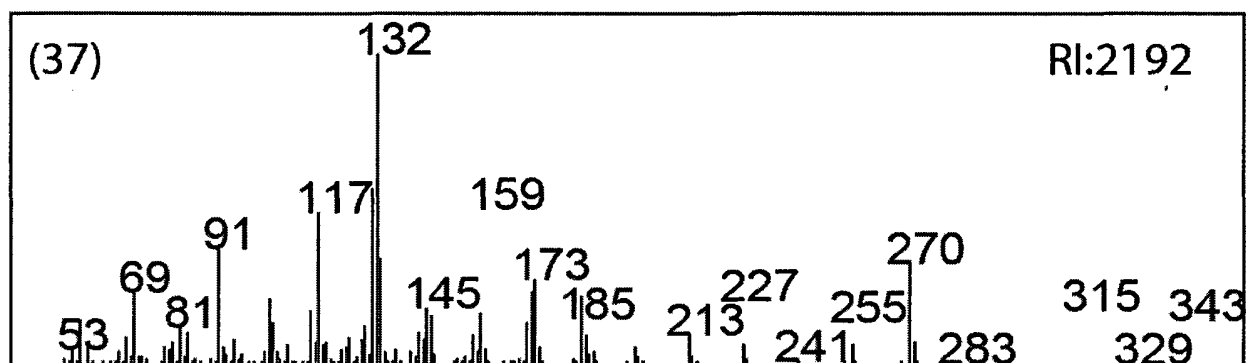
**Fig. 6 continued**

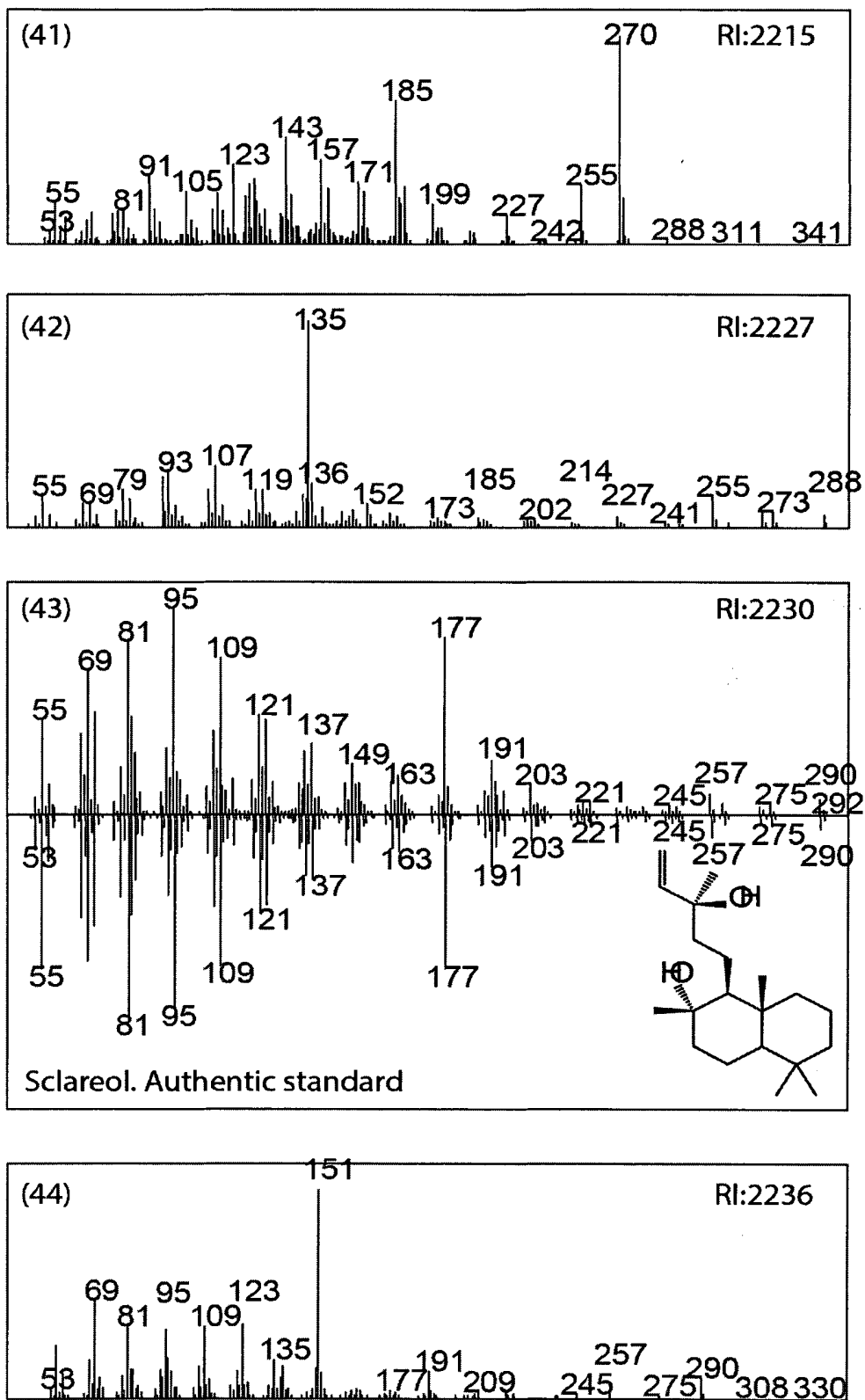
23 / 27

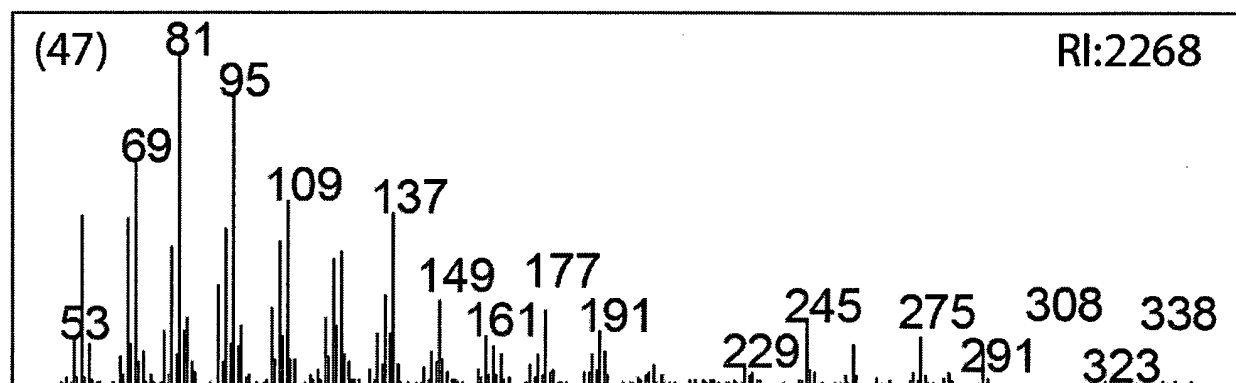
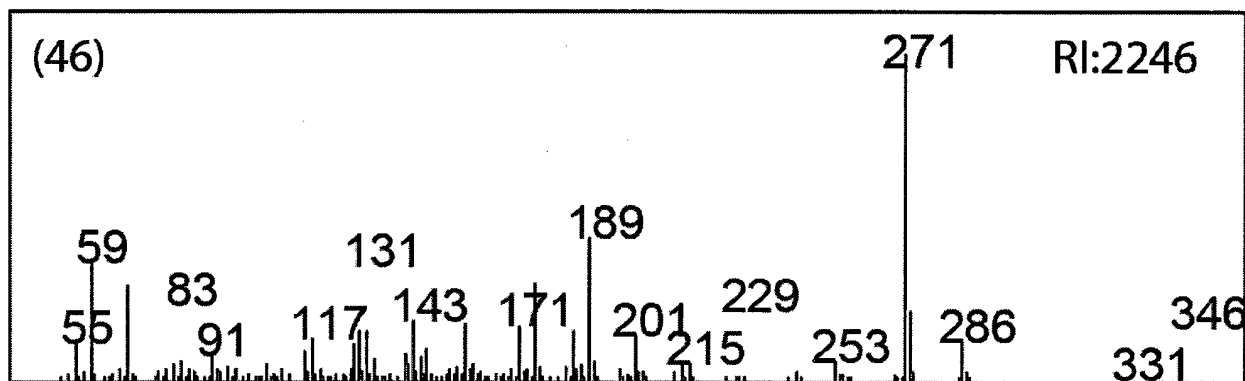
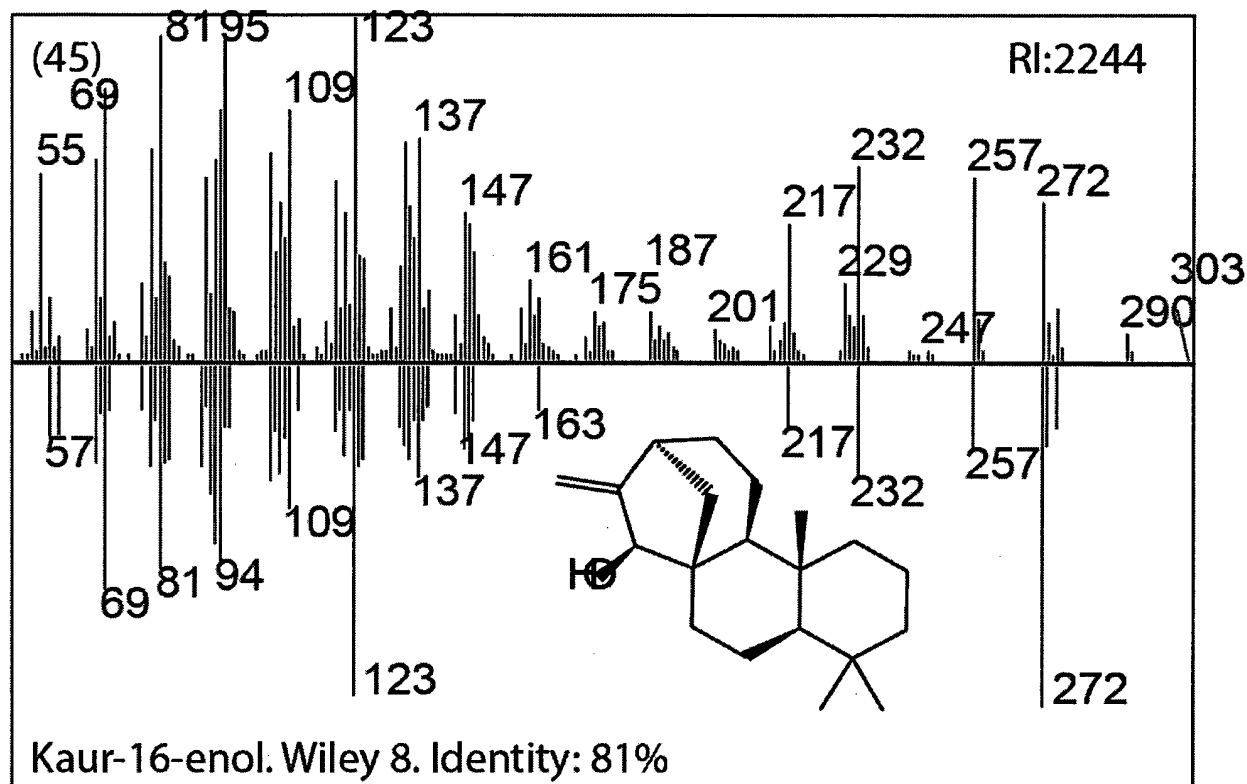
**Fig. 6 continued**

24 / 27

**Fig. 6 continued**

**Fig. 6 continued**

**Fig. 6 continued**

**Fig. 6 continued**

INTERNATIONAL SEARCH REPORT

International application No

PCT/DK2015/050021

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12P5/00 C12P7/00 C07K14/415 C12N9/00
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12P C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EP0-Internal, FSTA, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE UniProt [Online]</p> <p>18 September 2013 (2013-09-18), ZERBE ET AL: "Gene discovery of modular diterpene metabolism in nonmodel systems", XP002738980, Database accession no. R9UNP0 * See the sequence = 100% identity with SEQ ID NO 2 * ----- -/--</p>	16

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

28 April 2015

Date of mailing of the international search report

22/05/2015

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Korsner, Sven-Erik

INTERNATIONAL SEARCH REPORT

International application No

PCT/DK2015/050021

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE UniProt [Online]</p> <p>18 September 2013 (2013-09-18), ZERBE ET AL: "Gene discovery of modular diterpene metabolism in nonmodel systems", XP002738981, Database accession no. R9UPX6 * See the sequence = 100% identity with SEQ ID NO 15 *</p> <p>-----</p>	23
X	<p>DATABASE UniProt [Online]</p> <p>18 September 2013 (2013-09-18), ZERBE ET AL: "Gene discovery of modular diterpene metabolism in nonmodel systems", XP002738982, Database accession no. R9UM66 * See the sequence = 100% identity with SEQ ID NO 16 *</p> <p>-----</p>	24
Y	<p>ANDERSEN-RANBERG ET AL: "Expanding the molecular diversity through synthetic biology: Using combinatorial biochemistry for reconstruction of pathways to high-value and novel diterpenes", TERPNET 2013 (11th international meeting on biosynthesis, function and biotechnology of isoprenoids in terrestrial and marine organisms); Crete, Greece, June 2013 Book of Abstracts, 2013, page 93, XP002738983, Retrieved from the Internet: URL: <a href="http://terpnet2013.pharm.auth.gr/TERPN
 ET_BOOK_OF_ABSTRACTS.pdf">http://terpnet2013.pharm.auth.gr/TERPN ET_BOOK_OF_ABSTRACTS.pdf [retrieved on 2015-04-27] * See Abstract 058 (the oral disclosure may have been more detailed - to be settled) *</p> <p>-----</p> <p style="text-align: center;">-/--</p>	1-52

INTERNATIONAL SEARCH REPORT

International application No

PCT/DK2015/050021

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>PATERAKI ET AL: "Manoyl oxide as a precursor for forskolin biosynthesis: identification and characterization of the involved biosynthetic enzymes from Coleus forskohlii", TERPNET 2013 (11th international meeting on biosynthesis, function and biotechnology of isoprenoids in terrestrial and marine organisms); Crete, greece, June 2013 Book of Abstracts, 2013, page 60, XP002739006, Retrieved from the Internet: URL:http://www.plantengine.eu/sites/default/files/Book%20of%20abstracts%20020813.pdf [retrieved on 2015-04-27] * See Abstract 025 (the oral disclosure may have been more detailed - to be settled) *</p> <p>-----</p>	1-52
Y	<p>PATERAKI ET AL: "Manoyl oxide (13R), the biosynthetic precursor of forskolin, is synthesized in specialized root cork cells in Coleus forskohlii", PLANT PHYSIOLOGY, vol. 164, 30 January 2014 (2014-01-30), pages 1222-1236, XP002724033, * See page 1222 (Abstract), note the supplemental data; early online publication *</p> <p>-----</p>	1-52
Y	<p>MITCHELL: "Identification and characterization of diterpene synthases in the salvinorin A biosynthetic pathway", The University of Calgary Ph.D. thesis, 2012, pages 1-2, XP002738984, Retrieved from the Internet: URL:http://theses.ucalgary.ca/bitstream/11023/171/5/ucalgary_2012_mitchell_rod.pdf [retrieved on 2015-04-27] * See page 2 (Abstract); only the first two pages are cited here *</p> <p>-----</p>	1-52
Y	<p>ZERBE ET AL: "Gene discovery of modular diterpene metabolism in nonmodel systems", PLANT PHYSIOLOGY, vol. 162, 2013, pages 1073-1091, XP002724041, cited in the application * See page 1074 (Figure 1) and pages 1086-1087 (General applications and conclusion) *</p> <p>-----</p> <p style="text-align: center;">-/--</p>	1-52

INTERNATIONAL SEARCH REPORT

International application No

PCT/DK2015/050021

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ZERBE ET AL: "Bifunctional cis-abienol synthase from Abies balsamea discovered by transcriptome sequencing and its implications for diterpenoid fragrance production", THE JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 287, 2012, pages 12121-12131, XP055069746, * See pages 12121-12122 (Introduction, esp. Figure 1) *	1-52
Y	----- CANIARD ET AL: "Discovery and functional characterization of two diterpene synthases for sclareol biosynthesis in Salvia sclarea (L.) and their relevance for perfume manufacture", BMC PLANT BIOLOGY, vol. 12, 2012, pages 1-13, XP021128059, * See page 119 (Abstract) *	1-52
A	----- WO 2013/075239 A1 (THE UNIVERSITY OF BRITISH COLUMBIA) 30 May 2013 (2013-05-30) * See pages 33-35 (section B3) and pages 55-56 (section F1) *	1-52
X,P	----- DATABASE UniProt [Online] 11 June 2014 (2014-06-11), PATERAKI ET AL: "Manoyl oxide (13R), the biosynthetic precursor of forskolin, is synthesized in specialized root cork cells in Coleus forskohlii", XP002738987, Database accession no. X5A4D6 * See the sequence = 100% identity with SEQ ID NO 5 *	18
Y,P	----- ANDERSEN-RANBERG: "Identification and characterization of biosynthetic parts involved in plant diterpenoid biosyntheses", The University of Copenhagen Ph.D. thesis (reference abstract), 2014, page 1, XP055183341, Retrieved from the Internet: URL: http://research.ku.dk/search/?pure=en%2Fpublications%2Fidentification-and-characterization-of-biosynthetic-parts-involved-in-plant-diterpenoid-biosyntheses(9b672fe8-15a1-4685-8204-ce711d76b041).html [retrieved on 2015-04-27] * See the Abstract; the complete disclosure is deciding (the actual date of publication has not yet been established) *	1-52
	----- -/--	

INTERNATIONAL SEARCH REPORT

International application No

PCT/DK2015/050021

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y,P	<p>HANSEN ET AL: "Evolutionary cues from functional switching of two closely related class II diterpene synthases", 53rd annual meeting of the Phytochemical Society of North America; Raleigh, North Carolina, August 2013 Book of Abstract, August 2014 (2014-08), XP002739007, Retrieved from the Internet: URL:http://www.pсна-online.org/PSNA2014book.pdf [retrieved on 2015-04-27] * See TWTPS14 and TwTPS21 (compare Claims 19-20); the oral disclosure may have been more detailed - to be settled *</p> <p>-----</p>	1-52
Y,P	<p>IGNEA ET AL: "Reconstructing the chemical diversity of labdane-type diterpene biosynthesis in yeast", METABOLIC ENGINEERING, vol. 28, 10 December 2014 (2014-12-10), pages 91-103, XP002738988, * See page 92 (Figure 1); early online publication *</p> <p>-----</p>	1-52
Y,P	<p>GONG ET AL: "Diterpene synthases and their responsible cyclic natural products", NATURAL PRODUCTS AND BIOPROSPECTING, vol. 4, 18 April 2014 (2014-04-18), pages 59-72, XP002739034, * See page 59 (motifs); online publication *</p> <p>-----</p>	11,14

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/DK2015/050021

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2013075239 A1	30-05-2013	AU 2012343263 A1	05-06-2014
		CA 2855171 A1	30-05-2013
		EP 2783004 A1	01-10-2014
		US 2013224809 A1	29-08-2013
		US 2015037854 A1	05-02-2015
		WO 2013075239 A1	30-05-2013
